HEC MONTRÉAL

Essais sur l'adéquation cognitive entre l'état psychophysiologique d'un utilisateur et le format de l'information dans une interface de formation : le cas des instructeurs de vol en simulateur

Par

Christophe Lazure

Sciences de la gestion

(Option Technologies de l'information)

Sous la direction de

Pierre-Majorique Léger et Sylvain Sénécal

Mémoire présenté en vue de l'obtention du grade de

maîtrise ès sciences en gestion (M. Sc.)

Décembre 2018

Sommaire

Les instructeurs de vol accompagnent les pilotes tout au long de leur processus d'apprentissage et jouent un rôle crucial dans le processus de formation continue du domaine de l'aviation. Ainsi, pour améliorer la formation des pilotes, une compréhension des composantes pouvant influer sur la performance des instructeurs de vol est nécessaire. Ce mémoire par articles s'intéresse à l'état psychophysiologique des instructeurs de vol. Plus spécifiquement, ce mémoire cherche à comprendre l'expérience vécue par les instructeurs de vol dans un contexte de formation en simulateur. Lors des sessions de formation, les instructeurs de vol recourent à des interfaces qui agissent comme outils d'aide à la décision pour évaluer la performance des pilotes. Ce mémoire cherche également à comprendre comment les représentations de l'information peuvent influer sur la performance des instructeurs de vol. L'objectif de ce mémoire est de répondre aux questions suivantes : *Quels facteurs influencent l'état psychophysiologique des instructeurs de vol dans un contexte de formation en simulateur? Dans un contexte où des représentations dynamiques de l'information sont présentées simultanément à un utilisateur, est-ce qu'il y a des modèles de visualisation associés à de meilleures performances?*

En premier temps, une revue de littérature systématique a permis d'identifier 20 articles s'étant adressés directement aux facteurs de performance des instructeurs de vol depuis 1965. Par la suite, un examen de la portée (*scoping review*) a permis d'identifier une série de facteurs, reliés au domaine de l'interaction humain-machine, ayant un impact direct ou indirect sur les performances décisionnelles des instructeurs de vol dans un contexte de formation en simulateur. En second lieu, une étude en laboratoire intra-sujet menée auprès de 10 participants experts a permis de comprendre comment les représentations de l'information peuvent influer sur la performance des instructeurs de vol dans un contexte de formation en simulateur. L'analyse conjointe de mesures auto-déclarées et psychophysiologiques démontre qu'une adéquation cognitive entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation a un effet sur la performance psychophysiologique des instructeurs de vol au niveau du temps d'évaluation et de la charge cognitive.

Les résultats de ce mémoire contribuent à combler le manque dans la littérature sur les instructeurs de vol dans un contexte de formation en simulateur. Les parties prenantes de l'industrie de l'aéronautique peuvent utiliser les résultats de ce mémoire afin de mieux comprendre les facteurs de performance des instructeurs de vol et développer des interfaces mieux adaptées à leurs tâches.

Mots clés : Instructeurs de Vol, CFI, Représentation de L'Information, Charge cognitive, Attention visuelle, Adéquation cognitive, Cognitive Fit Theory, CFT.

Table des matières

Liste des Tableaux

Liste des figures

Liste des Abréviations, Sigles et Acronymes

CFI: « *Certified Flight Instructor* »

CFT: « *Cognitive Fit Theory* »

2D: « *Two-Dimensional* »

3D: « *Three-Dimensional* »

Avant-Propos

Ce mémoire a été rédigé par articles selon l'approbation de la direction du programme de la Maîtrise ès sciences en gestion en Technologies de l'Information. L'approbation des coauteurs a été obtenue pour chacun des deux articles se trouvant dans ce mémoire.

Le comité d'éthique de la recherche de HEC Montréal a donné son approbation pour cette expérience réalisée en mai 2018 (# certificat 2018-3034).

Le premier article est une revue de littérature présentant une revue systématique sur l'étude des instructeurs de vol ainsi qu'un examen de la portée (*scoping review*) sur l'identification des facteurs de performance des instructeurs de vol dans un contexte de formation en simulateur. Cet article a été soumis à la revue *The International Journal of Aerospace Psychology*.

Le deuxième article cherche à comprendre l'influence des représentations de l'information sur les performances des instructeurs de vol en contexte de formation en simulateur. L'article porte sur le développement d'un paradigme expérimental visant à comprendre dans quelles mesures une adéquation cognitive entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur. Cet article est actuellement en processus de révision finale en vue d'être soumis à la revue *The International Journal of Aerospace Psychology* après la complétion du mémoire.

Remerciements

La réalisation de ce mémoire n’aurait pas été possible sans la collaboration de nombreuses personnes exceptionnelles que je tiens particulièrement à remercier. Avant tout, je souhaite diriger mes premiers remerciements vers mes directeurs de recherche, Pierre-Majorique Léger et Sylvain Sénécal. C’est grâce à vos conseils et votre soutien que j’ai pu réaliser ce mémoire. Je vous remercie de m’avoir accordé votre confiance et de m’avoir soutenu tout au long de ce processus. Vous m’avez appris à structurer ma pensée selon un raisonnement scientifique, à comprendre les enjeux d’une recherche scientifique et à gérer les attentes de nos partenaires. Avec le Tech3Lab, vous avez su créer un environnement de travail dynamique, enrichissant et valorisant pour l’ensemble des étudiants sous votre supervision. Je serai à jamais reconnaissant d’avoir eu la chance d’effectuer mon mémoire au Tech3Lab.

Je tiens particulièrement à remercier la chercheuse postdoctorale assignée à mon projet, ma mentore, Laurence Dumont. Je n’aurais su réaliser ce mémoire sans ton aide, ton support et l’expérience que tu as apportés au projet. C’est en te côtoyant que j’ai eu la chance de développer des compétences qui me suivront tout au long de ma carrière. Ton professionnalisme et ta vivacité d’esprit ont été une source constante d’inspiration.

Je voudrais également remercier mes collègues du Tech3lab Sébastien Lourties, Benjamin Maunier, Robin Houenoussi et Tanguy Dargent pour votre amitié fraternelle. J’exprime également ma gratitude envers mes collègues et amis à la maîtrise Antoine, Marianne, Karl-David, Théophile, Yannick, Alexandre et Michael.

Je remercie également l’ensemble de l’équipe opérationnelle du Tech3lab, dont David, Bertrand, Beverly et Vanessa. Je remercie également Sophia et Arielle pour votre apport particulier au premier article de mon mémoire. Je tiens également à remercier spécialement Shang Lin de m’avoir aidé avec son expertise en statistiques. Merci également à tous les assistant(e)s ayant collaboré à notre expérience, plus particulièrement à Emma et Salima.

Je remercie également mon partenaire de recherche pour l’opportunité de ce projet ainsi que la Chaire de recherche industrielle CRSNG-Prompt en expérience utilisateur pour leur soutien financier.

Pour terminer, je tiens à remercier les membres de ma famille pour leur soutien inconditionnel, leurs encouragements et leur amour pendant toutes ces années universitaires. Finalement, merci à la meilleure compagne, Marie-Jeanne, de me motiver à toujours vouloir me dépasser. Je suis fier de pouvoir partager avec vous la réalisation de ce mémoire.

Chapitre 1 : Problématique et questions de recherche

Mise en contexte et justification de l'étude

L'industrie de l'aviation fait présentement face à une pénurie mondiale de pilotes d'avion. Selon une étude de Boeing (2018), 790 000 nouveaux pilotes de vol seront requis sur le marché mondial au cours des deux prochaines décennies. Cette demande impose une pression considérable sur le marché pour former rapidement de nouveaux pilotes. L'industrie doit optimiser le processus de formation des nouveaux pilotes afin de mieux s'adapter à cette pénurie de main-d'œuvre. C'est en misant, entre autres, sur les instructeurs de vol certifiés que l'industrie sera en mesure de faire face à cette pénurie. Les instructeurs de vol sont responsables de la prise en charge de la formation des pilotes. Ils accompagnent les pilotes à chaque étape de leur processus d'apprentissage et veillent à ce qu'ils possèdent les requis nécessaires pour obtenir leur licence. Comprendre les facteurs de performance des instructeurs pourrait aider à améliorer le processus de formation et donc, aider à répondre efficacement à la demande du marché. Dans ce mémoire, nous nous intéressons aux facteurs pouvant influer sur la performance des instructeurs de vol dans un contexte de formation en simulateur.

Plusieurs facteurs peuvent influencer la performance des instructeurs. Ces facteurs sont compris dans deux catégories : les compétences techniques et non techniques (Thomas (2000). Les compétences techniques font référence aux connaissances et savoirs acquis durant leur carrière de pilote. Au niveau des compétences non techniques, un cadre d'évaluation fourni par Thomas (2000) classifie ces compétences selon quatre grandes catégories : les processus communicationnels, la conscience situationnelle, la gestion des tâches et les processus éducationnels. Les processus éducationnels et communicationnels impliquent principalement l'interaction de l'instructeur avec les étudiants.

De nos jours, la formation des pilotes de vol se déroule majoritairement en simulateur (Socha *et al.*, 2016). La formation en simulateur permet de réduire les risques, améliorer la qualité des formations, réduire les coûts de formation et les coûts d'exploitation des aéronefs (Aragon and Hearst, 2005). Dans un contexte où la formation des pilotes de vol se déroule en simulateur, les facteurs de performance reliés à la conscience situationnelle et à la gestion des tâches doivent être davantage considérés. Durant une session de formation en simulateur, les instructeurs ont recours à différentes interfaces et sources d'informations pour évaluer les performances des pilotes (González Vega, 2002). Ces interfaces agissent comme outils d'aide à la décision lors de l'évaluation de la performance d'un pilote de vol. Les informations fournies par ces interfaces doivent être prises en compte afin d'évaluer leur impact sur le processus de décision des instructeurs de vol.

Pour favoriser le développement du pilote, les instructeurs de vol doivent être conscients des manœuvres effectuées par celui-ci, mais également des informations présentes sur l'interface. Une attention particulière aux différentes vues (ex. un instrument de vol, une vue en trois dimensions de l'aéronef, une vue graphique en deux dimensions de la trajectoire de l'aéronef) présentes sur leur interface est essentielle afin de maintenir l'engagement du pilote et de ne pas leur fournir d'informations ou de techniques erronées. Afin d'obtenir une vue globale des facteurs de performance des instructeurs, l'évaluation de l'interaction entre l'instructeur et les interfaces doit donc être également considérée. À notre connaissance, aucune étude ne s'est directement intéressée à ce sujet.

Les interfaces de pointe utilisent une combinaison d'instruments de vol, de visualisations en deux dimensions (ex. une représentation graphique de la trajectoire de l'aéronef) et en trois dimensions (ex. une vue en trois dimensions de l'aéronef) pour présenter les données de la leçon de vol. Les instruments de vol fournissent des données chiffrées d'une manœuvre de vol. Selon les standards en aviation (Transport Canada, 2016), un pilote doit constamment surveiller ses instruments pour évaluer les conditions de vol externe (ex. altitude, météo, visibilité, etc.) et les conditions de vol interne (ex. niveau de carburant, vitesse, avion, etc.) et effectuer une manœuvre en fonction de ces paramètres. Lors d'un entraînement en simulateur, les instructeurs doivent également avoir accès à une représentation de ces instruments de vol, car ils lui permettent d'évaluer si l'élève a pris la bonne décision selon les conditions du vol. Les instruments de vol sont essentiels aux instructeurs, car ils fournissent une vue en temps réel des paramètres ayant une incidence sur la décision du pilote.

Les représentations en deux dimensions ou en trois dimensions permettent, quant à elles, aux instructeurs de vol d'avoir des représentations apportant un visuel supplémentaire des manœuvres d'un pilote. Une représentation en deux dimensions permet d'apprécier graphiquement certains indicateurs tels que l'écart de vitesse par rapport aux standards, la trajectoire de l'avion ou l'écart de cette trajectoire par rapport aux standards. La représentation en trois dimensions fournit une vue de l'environnement et une appréciation du champ visuel du pilote. Les représentations en deux dimensions ou trois dimensions fournissent des indicateurs visuels qui peuvent faciliter l'évaluation de la performance du pilote.

Bien que le domaine de l'interaction humain-machine ait toujours été fortement ancré dans la recherche en aéronautique, la plupart des ouvrages antérieurs se sont concentrés sur les pilotes de vol (Landry, 2009). Le processus de traitement de l'information représente un aspect majeur au niveau des recherches en aviation et au niveau de la conception de systèmes aérospatiaux (Vidulich *et al.*, 2010; Peißl, Wickens and Baruah, 2018). Une meilleure compréhension du processus de traitement de l'information effectué par les instructeurs de vol est nécessaire pour améliorer l'efficacité et la cadence des formations.

Plusieurs études ont examiné l'impact des représentations sur les performances des pilotes de vol en étudiant les bénéfices et avantages relatifs des représentations en trois dimensions de basse fidélité et des représentations en deux dimensions (ex. Olmos et al., 2000; Wicken et al., 2000; Smallman *et al.*, 2001; Alexander, Wickens and Merwin, 2005). Bien que l''impacts de ces types représentations sur la performance dépend de la tâche à accomplir, des résultats non constants à travers les études n'ont su démontrer clairement quelle représentation possédait le plus d'avantages par rapport à l'autre lors d'une tâche d'évaluation de manœuvres aériennes, ce qui met en évidence une lacune dans la littérature. À notre connaissance, aucune étude ne s'est intéressée directement à l'influence des représentations de l'information sur les performances des instructeurs de vol. La théorie de l'adéquation cognitive permet de comprendre l'influencse des représentations de l'information sur la performance (Vessey, 1991). Cette théorie n'a jamais été appliquée au domaine de l'aviation pour comprendre l'impact des représentations sur la performance d'un utilisateur. (Vessey, 2006; Bacic, 2014).

Selon l'étude fondatrice de la théorie (Vessey, 1991), il existe une relation entre le format utilisé pour présenter l'information et le type de tâche à accomplir. Le format utilisé pour représenter le problème doit être aligné avec la stratégie décisionnelle de l'utilisateur (Vessey and Galletta, 1991; John and Kundisch, 2015). L'adéquation cognitive se réalise entre le format de la présentation et le type de la tâche lorsque le format de la présentation utilisé supporte les stratégies nécessaires à l'accomplissement de la tâche. La théorie stipule que lorsqu'une adéquation cognitive est atteinte entre la représentation d'un problème et une tâche, l'efficience et l'efficacité de la tâche sont améliorées (Vessey and Galletta, 1991).

Le développement des outils psychophysiologiques permet de mesurer d'une manière non intrusive et précise, les réactions implicites d'un utilisateur (de Guinea, Titah and Léger, 2014). Ces outils peuvent être utilisés pour évaluer l'adéquation cognitive (Bacic, 2014). Grâce à ces données psychophysiologiques, les effets de l'interaction sur la charge cognitive d'un utilisateur peuvent être mesurés. Puisque la charge cognitive (Sweller, Ayres and Kalyuga, 2011) engendre des effets sur la performance, ce mémoire propose de considérer ce construit pour évaluer l'impact des représentations sur la performance des instructeurs. En se basant sur la théorie de l'adéquation cognitive et grâce à l'utilisation de ces outils, ce mémoire a permis d'observer dans quelles mesures une adéquation cognitive entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur.

Objectifs de l'étude et questions de recherche

Le premier objectif de ce mémoire est de faire l'état des connaissances académiques sur les instructeurs de vol. En nous basant sur l'état de la littérature, nous voulons proposer des pistes de recherche sur les facteurs potentiels à surveiller et à prendre en compte dans l'évaluation de la performance des instructeurs de vol. Un premier article fournit une revue systématique de l'état de la littérature sur les instructeurs de vol et un examen de portée (*scoping review*) proposant une liste de facteurs à prendre en considération lors de l'évaluation de la performance des instructeurs de vol. Afin d'établir les facteurs qui entrent en jeu dans un contexte de formation en simulateur, nous allons élargir notre recherche aux sciences du comportement humain et à l'interaction humain-machine. De fait, ce mémoire pose la question suivante :

Question 1 : *Quels facteurs influencent l'état psychophysiologique des instructeurs de vol dans un contexte de formation en simulateur?*

Le deuxième objectif de ce mémoire est de comprendre l'influence de l'interaction humain-machine auprès des instructeurs de vol dans un contexte de formation en simulateur. Plus précisément, en se basant sur la théorie de l'adéquation cognitive, ce mémoire tente de comprendre l'effet des représentations dynamiques de l'information sur la performance des instructeurs de vol. Ce mémoire utilise différents modèles de visualisation pour évaluer l'influence des représentations dynamiques de l'information sur la performance. Pour ce faire, ce mémoire a mené au développement d'un paradigme expérimental évaluant l'effet des représentations dynamiques de l'information, présentées simultanément à un utilisateur, sur la performance. De fait, ce mémoire pose la question suivante :

Question 2 : *Dans un contexte où les représentations de l'information sont présentées simultanément à un utilisateur, est-ce qu'il y a des modèles de visualisation associés à de meilleures performances.*

Pour démontrer le potentiel et la faisabilité de ce paradigme expérimentale, une étude en laboratoire a été réalisée auprès de 10 participants experts. Les participants ont dû évaluer un pilote effectuant une série de manœuvres d'atterrissage et de décollage.

Contribution potentielle

D'un point de vue théorique, ce mémoire fournit plusieurs contributions. Premièrement, ce mémoire fait le point sur l'état de la connaissance académique sur les instructeurs de vol. Une revue systématique sur les instructeurs de vol et un examen de la portée (*scoping review*) sur les facteurs de performance des instructeurs de vol dans un contexte de formation en simulateur ont donc été réalisés. Ce mémoire fournit une liste de concepts pouvant être étudiés lors d'une étude future. L'auteur de ce mémoire espère que ces revues de littérature permettront d'augmenter la compréhension des facteurs ayant une influence sur les performances des instructeurs de vol et ainsi, combler les lacunes présentes dans la littérature sur les instructeurs de vol dans un contexte de formation en simulateur.

Deuxièmement, ce mémoire espère également développer un paradigme expérimental, basé sur le modèle de l'adéquation cognitive permettant de déterminer dans quelle mesure une adéquation cognitive entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur. Ce mémoire contribue à la littérature en étendant la théorie de l'adéquation cognitive à de nouvelles dimensions. Ce mémoire utilise le modèle de l'adéquation cognitive afin d'analyser l'effet des représentations dynamiques de l'information sur les performances des utilisateurs. Ce mémoire développe un paradigme qui pourra être utilisé par d'autres chercheurs, pour évaluer l'impact des représentations dynamiques de l'information sur les performances des utilisateurs.

D'un point de vue pratique, ce mémoire fournit plusieurs implications. Premièrement, l'identification des facteurs de performance illustrés dans ce mémoire pourrait permettre à la communauté aéronautique de développer de meilleures pratiques en matière de sélection, de formation et de développement des instructeurs.

Deuxièmement, le développement du paradigme expérimental énoncé précédemment pourrait permettre aux spécialistes de l'aviation de développer des interfaces mieux adaptées aux tâches des instructeurs de vol. Les praticiens pourront utiliser ce paradigme afin de tester la conception et le design des interfaces. De plus, l'utilisation de modèles innovants de visualisation pourrait permettre aux spécialistes des facteurs humains de concevoir des interfaces s'adaptant à l'attention visuelle des utilisateurs.

Structure du mémoire

Ce mémoire est structuré sous la forme de deux articles. Il s'adresse tant aux spécialistes des facteurs humains œuvrant dans le domaine de l'aéronautique qu'aux chercheurs en technologies de l'information et aux professionnels en expérience utilisateur. Le premier article permet de faire l'état des connaissances académiques sur les instructeurs de vol et les construits attribuables à des domaines connexes pouvant informer sur les facteurs de performance des instructeurs de vol. Le second article étudie comment les représentations dynamiques de l'information peuvent influencer certains des construits identifiées dans le premier article et donc, avoir un impact sur les performances des instructeurs de vol. Pour ce faire, l'article développe un paradigme expérimental qui permet d'étudier dans quelles mesures une adéquation cognitive, entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur. Ces articles permettent l'obtention d'une meilleure compréhension des facteurs pouvant affecter les performances des instructeurs de vol en contexte d'entraînement en simulateur.

Informations sur l'article 1

Le premier article a été soumis à la revue *The International Journal of Aerospace Psychology*. Cette revue s'adresse aux professionnels de l'aviation, mais également aux chercheurs de domaines connexes tels que l'informatique qui s'intéressent au développement et à la gestion de systèmes aérospatiaux. L'article est présentement en révision auprès de l'éditeur. Cet article a pour objectif de recenser les études qui se sont directement intéressées aux instructeurs de vol et d'établir une liste de concepts provenant de la psychologie cognitive et de la psychophysiologie permettant d'expliquer les composantes influençant la performance des instructeurs de vol. Cet article a permis d'identifier des concepts vers lesquels nous avons orienté notre recherche pour le deuxième article.

Résumé de l'article 1

L'article présente une revue de littérature sur les instructeurs de vol. La revue de littérature systématique portant sur l'étude des instructeurs de vol dans un contexte de formation en simulateur, permet d'identifier seulement 20 articles s'étant directement intéressés aux instructeurs de vol depuis 1965. De ces 20 articles, deux d'entre eux ont été publiés avant les années 2000, dix entre les années 2000 et 2010 et huit après l'année 2010. Même si le nombre d'articles publiés sur les instructeurs de vol est limité, le fait que 18 articles sur 20 aient été écrits au cours des deux dernières décennies montre que ce sujet retient de plus en plus l'attention de la communauté académique. En utilisant la même classification des facteurs de performance que le cadre proposé par Thomas (2000), ces études peuvent être classées en quatre catégories de compétences non techniques : processus éducatifs (6);

communication (5); gestion des tâches (4); conscience de la situation (1). Les trois autres études peuvent être classées dans une catégorie de compétences techniques. Le dernier des 20 articles est celui de Thomas (2000). Une quantité limitée d'articles a été réalisée sur la gestion des tâches et la conscience situationnelle. Des compétences dans lesquelles les outils de mesures psychophysiologiques pourraient fournir une mesure claire du processus attentionnel et décisionnel des instructeurs de vol. Une seule étude a eu recours à des outils de mesures psychophysiologiques (Sohn and Jo, 2003). Cette étude a utilisé les outils de mesures psychophysiologiques afin d'évaluer la charge mentale des pilotes et non celle des instructeurs de vol. Cela démontre une lacune dans la littérature, dû au nombre d'articles limités ayant utilisé les outils de mesures psychophysiologiques pour évaluer les instructeurs de vol.

Afin d'augmenter notre apprentissage sur le sujet, un examen de la portée (*scoping review*) portant sur l'identification des facteurs de performance des instructeurs de vol a été réalisé. Cette revue s'intéresse principalement aux facteurs de performance reliés à la conscience situationnelle. Les construits suivants ont été identifiés : la conscience situationnelle, l'attention, la détection des changements, la tolérance aux risques, le flow et l'absorption cognitive, la charge cognitive et le stress. La sélection de ces concepts est basée sur le modèle de Thomas (2000) et sur les sciences du comportement humain. Pour chaque construit, l'article fournit plusieurs éléments : (1) une définition du construit; (2) une explication de la pertinence de considérer ce construit pour évaluer la performance des instructeurs de vol; (3) les mesures psychophysiologiques et psychométriques de ce construit; (4) l'application de ce construit à des domaines connexeInformations sur l'article 2

Le deuxième article est présentement en préparation en vue d'être soumis à la revue *The International Journal of Aerospace Psychology* après la complétion du mémoire. Le deuxième article cherche à comprendre l'influence des représentations dynamiques de l'information sur les performances des instructeurs de vol dans un contexte de formation en simulateur. L'article a pour objectif de développer un paradigme expérimental permettant d'évaluer dans quelles mesures une adéquation cognitive, entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur. Ce paradigme utilise différents modèles de visualisation pour évaluer l'influence des représentations dynamiques de l'information sur la performance.

Résumé de l'article 2

Se basant sur la théorie de l'adéquation cognitive (Vessey and Galletta, 1991), ce deuxième article étudie l'impact des représentations dynamiques de l'information sur la prise de décision des instructeurs de vol dans un contexte de formation en simulateur. Cet article développe un paradigme expérimental permettant d'analyser l'effet des représentations dynamiques de l'information lorsque celles-ci sont présentées simultanément à un utilisateur. Afin de tester la faisabilité de ce paradigme, une étude en laboratoire a été réalisée auprès de 10 participants experts dans le but de récolter des mesures auto-rapportées (explicites) ainsi que des mesures comportementales et biométriques (implicites). Analysant dynamiquement l'adéquation cognitive à l'aide de l'ordre de la séquence visuelle des participants, les résultats de cette étude suggèrent que différents modèles de visualisation ont une influence sur les performances des instructeurs de vol au niveau du temps d'évaluation et de la charge cognitive. Le développement de ce paradigme expérimental fournit aux professionnels de l'aviation une nouvelle méthode pour évaluer l'effet des représentations dynamiques de l'information sur les utilisateurs et potentiellement développer des interfaces mieux adaptées à leur tâche.

Contributions aux articles de recherche

Afin de mieux comprendre mon apport aux deux articles, le tableau ci-dessous présente les étapes du processus de recherche et ma contribution pour chacune d'elles. J'inclus le pourcentage du travail que j'ai effectué à chacune de ces étapes.

Étapes du processus	Contribution
Définition des requis du partenaire	Gestion des attentes du partenaire **– 50%** - Participation aux rencontres hebdomadaires avec le partenaire pour établir les requis du projet, les objectifs de recherches et faire la gestion du projet (février à avril). - Définir les questions de recherche dans les articles
Revue de littérature	Effectuer la revue de littérature pour déterminer les facteurs de performance des instructeurs de vol et les construits testés pour étudier l'influence des représentations de l'information sur la performance **– 90%** Définir les outils de mesure utilisés pour tester les construits – **75%**

Stimulus	Production et extraction des stimulus **– N/A** (Les stimulus ont été produits et extraits par le partenaire.) Inspection de la validité écologique des stimulus **– 50%** Choix des stimulus pour l'ensemble des tâches **– 50%**
Design expérimental	Élaborer la demande au CER et des demandes de changement au dossier **– 50%** - Documenter tout changement dans le design expérimental et le reporter au CER - L'équipe opérationnelle du Tech3lab a contribué à développer les formulaires de consentement et de compensation Concevoir le protocole d'expérimentation **– 50%** Mener des cycles de prétest afin de peaufiner le design expérimental : Outils de mesure utilisés, enchainement des tâches, choix des stimulus **– 100%** Mettre en place la salle de collecte **– 50%**
Recrutement Prétests	Recrutement des participants : solliciter, contacter, organiser la participation – **80%** (Le partenaire s'est chargé de faire parvenir notre lettre de sollicitation aux participants potentiels) Le partenaire était responsable des compensations.
Prétests et collectes	Responsable des opérations lors de la collecte **– 60%** Participation à l'ensemble des collectes : assistance technique et aide aux assistantes pour tout problème avec la salle de collecte **– 100%**
Extraction et transformation des données	Extraction et mise en forme des données physiologiques, psychométriques, cognitives et émotionnelles pour permettre l'analyse statistique **– 100%**
Analyse de données	Analyse des données psychophysiologiques **– 100%** (lecteur d'émotions faciales, senseurs d'activité électrodermale) : une formation effectuée par le Tech3lab m'a permis de mener de manière autonome cette analyse. Analyses statistiques **– 75%** - Aide sur SAS pour les analyses par le statisticien de la Chaire de recherche industrielle CRSNG-Prompt en expérience utilisateur

Rédaction	Contribution dans l'écriture des articles – **90%** - Les autres auteurs ont apporté des commentaires constructifs qui m'ont permis de peaufiner les articles

Chapitre 2 : Article 1

Evaluation of certified flight instructors' performance factors in a flight simulation training context

Christophe Lazure[1], Laurence Dumont[1], Sofia El Mouderrib[2], Jean-François Delisle[3], Sylvain Sénécal[1] et Pierre-Majorique Léger[1]

[1]HEC Montréal, Montréal, QC H3T 2A7, Canada

{christophe.lazure,laurence.dumont,sylvain.senecal,pierre-majorique.leger}@hec.ca

[2]UQAM, Montréal, QC H3C 3P8, Canada

el_mouderrib.sofia@courrier.uqam.ca

[3]CAE Inc., Montréal, QC H4T 1G6, Canada

jean-francois.delisle@cae.com

Abstract

A. Objective: We conducted a systematic review of peer-reviewed articles aimed at the evaluation of certified flight instructors' (CFI) performance in a flight simulation training context and a scoping review of potential research avenues given the previously identified gaps.

B. Background: As the demand for pilots will continue to grow significantly in the coming decades, so will the demand of CFIs, and ways to improve their existing performance. Understanding performance factors of CFIs could benefit their training and help meet the increasing demand for pilots.

C. Method: State-of-the-art research on the subject was surveyed via a systematic review of performance factors of CFIs and a scoping review to identify areas where other fields of research could inform CFI performance evaluation.

D. Result: Only 20 articles since 1965 have directly assessed performance factors of CFIs. Their focus has mostly been on communication and educational processes. The scoping review brings forward concepts from cognitive psychology and psychophysiology as means of improving the current understanding of CFI situation awareness and task management.

E. Conclusion: Very little work has been done on CFI situation awareness and task management. These are the two main domains in which psychophysiological tools could provide a clear understanding of the attentional and decisional processes at play while developing situation awareness in a dynamic environment and quantify the task load affecting it.

Keywords: certified flight instructors; flight instructors; flight training; situation awareness; task management

Introduction

With between 255,000 and 637,000 new pilots required on the market in the next two decades (CAE, 2017; The Boeing Company, 2017), airlines are under tremendous pressure to train new pilots faster, without either compromising on safety or flouting national and international standards. Certified flight instructors (CFI) accompany pilots during every step of their learning process and play a crucial part in the instructional and continuous training process in aviation. Understanding the behaviors and characteristics of expert CFIs could help improve and individualize training to reach maximum performance potential, so as to provide each student with an optimal learning experience.

The first objective of this article is to identify the state-of-the-art in performance evaluation to establish a hypothesis on the potential factors that need to be monitored and acted upon. A systematic review will provide a list of the academic research that has been conducted while a scoping review will explore potential CFI performance factors.

For the purposes of putting the results of this systematic review into perspective, Thomas and colleagues (2000) provided a systematized framework to evaluate the work of CFIs. This framework lists the five following areas of performance: technical skills, educational processes, communication, situation awareness and task management. The aim was to be exhaustive on the themes present in the literature.

Technical skills refer to the knowledge and procedures acquired throughout their pilot career. Educational processes and communication mainly involve their interactions with students. The simulation training context increases the complexity of the course, since it relies on much more than just communication and educational processes. In a flight training simulator, a CFI must also be alert of the virtual environment, the data provided by his interface and the simulation environment. Situation awareness and task management come into play in a context where CFIs use different interfaces and sources of information in a training session to evaluate pilot performance (González Vega, 2002), and they need to take into account a vast amount of data and make an informed decision according to the information on the interface. In a training situation, CFIs must be aware of the training context, the actions taken by the pilot and the information on the interface to promote the success of the pilot's learning process. They need to focus on such information in order to keep their students engaged and to not provide them with erroneous information or techniques. The evaluation of a CFI’s performance is based on the following elements: appraisal of the training course, successful acquisition of knowledge and skills by the pilot, the CFI’s coaching ability and his interaction with the student. However, in order to have a holistic view of his performance, we also need to consider the evaluation of his interaction with the interface that he is using in a training simulation, a subject that, to our knowledge, has not been studied in the scientific literature.

To guide the scoping review, the disciplines related to human behavior sciences and human-computer interaction might broaden our knowledge on the study of CFI behavior in a flight simulation training context. Since situation awareness (SA) is a major concept in human-computer interaction and strongly correlated to decision-making (Ji et al., 2011), we conducted a scoping review that explored the factors that have a direct or indirect impact on SA. Such an exercise illustrates which factors could be considered to extend the literature that the first article has demonstrated.

Systematic Review of CFI Performance Factors

In this first section, we systematically examine peer-reviewed studies specifically related to CFIs. This will provide a better understanding of the components of flight instruction that have been experimentally studied and, more importantly, the areas of this field which still remain unappreciated.

Review Methods

A systematic review of papers that have identified performance factors of CFIs will provide a historic and state-of-the-art view of the field and help identify starting points for further investigation. In the first section of this article, we chose to use a qualitative systematic review for studies that directly address the performance of flight instructors, as this type of review is well suited for assessing gaps in a scientific literature (King and He, 2005; Paré et al., 2015).

The keywords used were Flight and Instructor, in two databases: Web of Science (Thomson Reuters Institute of Scientific Information) and Scopus (Elsevier). The selection of these databases provided the best complementarity between scientific fields (Guz and Rushchitsky, 2009) and time range (Aghaei Chadegani et al., 2013). Since the aviation field has been of interest for multiple disciplines (e.g. psychology, physiology, education, human factors, ergonomics) for several decades, this database combination best suited this systematic review.

Each abstract was subsequently screened to see if it met the two following criteria: (i) CFI performance must be the primary aim of the study in a way that provides direct insight into the role of CFIs or that can lead to applications in their training or work; (ii) CFIs should not be studied as expert pilots but as teaching professionals per se.

Results

Web of Science generated 148 results and Scopus 308, with an overlap of 83 articles. Three supplementary articles were fortuitously found during other literature searches conducted in the research leading to this manuscript, all of them on Google Scholar. One was published in a francophone peer-reviewed journal and two in a locally published peer-reviewed university paper. Although the databases cover other languages and local journals and theses, they could not be found on Scopus or on Web of Science. Hence, the abstracts of 370 articles were evaluated based on the aforementioned criteria.

A total of 20 studies satisfied the two criteria and were thus retained for this review. *Table 1* lists these papers along with an overview of the measures used and the performance variable that was studied.

Table 1. Articles retained for the qualitative systematic review on CFI performance

Study #	Author(s)	Country	Type	Measures	Performance variable
1	Bjerke & Malott (2011)	USA	Quantitative	Bespoke survey (not validated)	Cross-country flight experience
2	Byrnes (2017)	USA	Quantitative	Bespoke experimental simulation setup Bespoke survey (not validated)	Virtual learning environment
3	Harbeck, Kirschner, Wulle, & Bowen (2014)	USA	Quantitative	Bespoke survey (validated)	Experience
4	McDale & Ma (2008)	USA	Quantitative	Bespoke survey (validated)	Fatigue
5	Nählinder & Berggren (2005)	Sweden	Quantitative	Bespoke survey (not validated)	Embedded pedagogical tools

6	Gontar & Hoermann (2015)	Germany	Quantitative	NTS rating tools Line Operations Safety Audit (LOSA) (validated)	Within-group agreement and interrater reliability
7	Diels (2007)	USA	Quantitative	Defining Issues Test 2 (DIT2) (validated) 2 Bespoke items (not validated)	Ethics
8	Guion & Ikomi (2000)	USA	Quantitative	Judge’s Experience Questionnaire (JEQ) (validated) Bespoke interview grid (not validated)	Self-insight
9	Kreienkamp & Luessenheide (1985)	USA	Quantitative	Myers-Briggs Type Indicator (validated)	Personality
10	Sohn & Jo (2003)	Korea	Quantitative	Myers-Briggs Type Indicator (validated) Heart rate Altitude deviation NASA Task Load Index (validated) Questionnaire of the personality harmony (not validated)	Personality combination
11	Pomarolli & Ambler (1965)	USA	Quantitative	Instructor's voluntary withdrawal data	Voluntary withdrawals

12	Xu, Sohoni, McCleery, & Bailey (2006)	USA	Quantitative	Instructor schedule data computation	CFI scheduling problem
13	Burian & Feldman (2009)	USA	Quantitative / qualitative	Bespoke survey (not validated)	Mastery of weather content
14	Crow, Niemczyk, Andrews, & Fitzgerald (2011)	USA	Qualitative	N/A	Method in which flight instructors are trained
15	Hall (2011)	USA	Qualitative	Bespoke survey (open-ended questions)	Challenging student behaviors
16	Hale & Breaux (2011)	USA	Theoretical	N/A	Communication training
17	Hoover (2008)	USA	Theoretical	N/A	Educational learning theory
18	Lépinard (2014)	France	Theoretical	N/A	Communication
19	Thomas (2000)	Australia	Theoretical	N/A	Systematic method for the evaluation of instructor and student performance
20	Wetmore & Lu (2007)	USA	Theoretical	N/A	Style of conflict management

Interestingly, the only study that used physiological measurement tools (Sohn and Jo, 2003) used them to measure the pilot students' mental workload, and not the instructors', in order to correlate it with the student-instructor harmony in terms of personality.

A common feature is the use of surveys (11 studies) to investigate the knowledge, behavior, personality and cognition of flight instructors. Seven studies used at least one survey that was not validated with respect to internal coherence or test-retest fidelity. Nine studies used bespoke surveys, created by the research team, in their assessment.

Table 2. Classification of CFI performance factors

Skillset	Performance factor
Technical Skills	Cross-country flight experience (1)
	Experience (3)
	Scheduling problem (12)
Situation Awareness	Mastery of weather content (13)
Educational Processes	Virtual learning environment (2)
	Embedded pedagogical tools (5)
	Method in which flight instructors are trained (14)
	Communication training (16)
	Educational learning theory (17)
	Within-group agreement and interrater reliability (6)
Task Management	Fatigue (4)
	Ethics (7)
	Self-insight (8)
	Personality (9)
Communication	Challenging student behavior (15)
(between student and CFI)	Style of conflict management (20)
	Personality combination (10)
	Voluntary withdrawal from students (11)
	Conceptual tool for communication (18)

Table 2 contains a classification of performance factors using a classification for CFI non-technical skills built by Thomas (2000). The two categories that have elicited the most interest from research so far are Educational Processes and Communication, with six and five articles,

respectively. Interestingly, these two categories directly involve interaction with pilots. Task Management and Technical Skills have received slightly lower interest, with four and three studies each. Strikingly, very little research has been conducted on the Situation Awareness of CFIs, with only one study which didn't use validated questionnaires.

Discussion of the systematic review

This systematic review of peer-reviewed articles that directly measure CFI performance shows that the corpus of study on flight instructors is not very extensive. It consists of a handful of studies (20 articles) based on different theoretical backgrounds for which the topics range from personality assessment to ethics and physiology. This indicates a need to nurture this field of study in order to improve the state of knowledge regarding flight instructors' performance.

In this limited amount of papers, we found an uneven distribution of research topics which illustrates the need to investigate CFI characteristics in a more standardized manner. Using standardized and validated tools, or taking the necessary steps to validate bespoke surveys, will also generate more conclusive results and reliable tools that can be used by other researchers to avoid reproducibility issues in this emerging field of study.

Results suggest that the concept of human-computer interaction was not taken into account in the evaluation of CFI performance in a flight simulation training context. This field could improve our understanding of performance using broadly recognized concepts such as situation awareness and task management. In the aviation domain, training sessions are generally executed in a simulation training context which requires the CFI, to constantly analyze information on his interface to evaluate his trainee's performance.

Furthermore, given the importance of situation awareness in decision-making theories, research efforts in that direction would benefit from the quantity of previous research on the topic in aviation-related domains.

Scoping review of CFI performance through situational awareness factors

The results of the systematic review guided us towards focusing our attention on research concerning situation awareness (SA) and related constructs which could significantly improve knowledge on CFI performance in a training simulation context.

In this section, a scoping review was conducted to "map rapidly the key concepts underpinning a research area" (Arksey and O'Malley, 2005). This review was carried out on situation awareness and related constructs that are linked to CFI performance in a simulation training context. Given the many definitions and operationalization of SA in the literature, which strongly intersect with many other domains of study of human performance, we extended our focus to concepts and constructs that can be borrowed from many disciplines, namely cognitive psychology, sport psychology, neuroinformation systems and neuroleadership and psychophysiology. The selection of these concepts and constructs is partially based on Thomas' (2000) model, which classified a set of behavioral markers indicative of performance, and partially on the theoretical background provided by human behavior disciplines.

Given that the results of the systematic review show an infrequent use of validated psychometric tools and a quasi-absent use of psychophysiological tools, state-of-the-art measures will be brought forward for each concept.

Each section of this review provides us with cases highlighting the relevance of each concept to the CFI, the theoretical definition, the measures of the constructs and the application of the constructs in the related field. This section presents a holistic view of the literature that could fill the gap shown by the results of the systematic review and extend our knowledge in these areas.

Situation Awareness

As identified in the systematic review, only one study was interested in CFI situation awareness, which highlights the lack of knowledge in this field. Situation awareness is a classical concept and model in the aviation and military domain and can be defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 2015, p. 36). There are three levels of situation awareness, namely, perception, comprehension and projection (Endsley, 2015).

In aviation, SA can be defined as the pilot's mental model of the current state of the flight and mission environment (Hicks et al., 2014). In the context of the CFI's tasks, he could notice a maneuver by the pilot (perception), detect that this maneuver does not meet the safety standards (comprehension) and finally, understand that he will have to make the pilot practice this maneuver and test him on aviation safety standards (projection).

Situation awareness levels are usually measured using SART or CLSA (Hicks et al., 2014). The situation awareness rating technique (SART) was first introduced in 1990 by R.M. Taylor (Taylor, 1990) to develop a self-rating technique of SA (Endsley, 2015). China Lake Situational Awareness (CLSA) is a five-degree scale "that pilots use to rate their perceived level of SA for a mission or mission segment" (Hicks et al., 2014). O'Brien and O'Hare (2007) also demonstrated that a test of situational awareness ability, like WOMBAT, may be a predictive factor of a pilot training performance. The WOMBAT test is conducted to evaluate SA ability variables of a participant to "search, evaluate, and integrate information about all relevant events, conditions, and resources, quickly assess changes in situational priorities, and allocate attention accordingly" (Corl and LaRoche, 2001). A test of situational awareness ability measures various variables of SA such as perception, memory, attention and executive control.

Attention

As stated in the first level of the SA theory, pilots need to be fully aware of their environment, and attention is critical for monitoring and supervising the aircraft system (Wickens, Gempler and Ephimia, 2000). It is well-rooted in the theory that "attentional control is essential for successful performance in aviation" (Talleur and Wickens, 2003; Gibb, Gray and Scharff, 2010). Hence, attention is the first non-aviation psychology concept that this scoping review addresses.

Three divisions of visual attention, namely, oriented attention, divided attention and selective attention were shown to be in relation with flight performance training while the sustained attention test demonstrated a lack of correlation between sustained attention and flight performance after a training (Gray, Gaska and Winterbottom, 2016). Sustained attention refers to the ability to focus on a specific task during a period and not get distracted (O'Donnell, Moise and Schmidt, 2005). Selective attention refers to the ability to focus our attention on only one factor or stimulus and ignore other distractions (Ball and Owsley, 1993). Oriented attention is the ability to voluntarily shift attention from and to a specific location point between cognitive tasks (Posner, 1980). Divided attention, also known as multi-tasking, refers to the ability to focus our attention on multiple locations or tasks simultaneously (Ball and Owsley, 1993). Gray and colleagues (2016) also concluded that, especially in a landing test, divided attention was the most predictable attention value in a pilot evaluation.

When following the mind-eye hypothesis, a strong proxy of the location of overt visual attention can be found in eye-tracking (Van Orden, Jung and Makeig, 2000). Eye-tracking studies have revealed that sports experts primarily seek important information in specific areas and tend to perform more appropriate and efficient visual research than novices (Mann et al., 2007).

The importance of the three types of attention in CFIs could be assessed using standard neuropsychological tests, and the use of eye-tracking as a proxy for the location of visual attention could be a strong basis for further understanding the cognitive processes involved in the CFI's tasks.

Change detection

The CFI's task in maintaining good situation awareness includes the detection of many mistakes and changes both in the behavior of the pilot and in the interface used to monitor the state of the aircraft. The CFI should be responsive to stimuli and to the learning process of his student, and he should monitor the progress in the flight scenario and the actions of the pilot in order to compare this information with existing standards. Hence, the CFI's ability to effectively detect mismatches between expectations and the state of the simulation aircraft environment is crucial to situational awareness performance.

There are two main sources of failure in the detection of changes in one's surroundings: inattentional blindness and change blindness. Inattentional blindness occurs when an individual doesn't detect the occurrence of an unexpected event in his surroundings and change blindness is the absence of conscious acknowledgement of a change even though all the relevant information to detect such a change has been processed (Gibbs, Davies and Chou, 2016). In experimental settings, such phenomena are mostly studied through the flicker paradigm and, in a more ecological way, either through videos, simulations or real-life situations. These changes can be of any nature, but most studies (Gibbs, Davies and Chou, 2016) have focused on changes in the visual environment.

The flicker paradigm was developed by Rensink, Regan and Clark (1997) and it typically consists of quickly flashing two different variations of the same picture with a visual mask between the pictures (usually a blank screen) in a loop, until the participant either notices the change or 60 seconds have elapsed. Changes in the picture variations can either be central or peripheral, where central changes are defined as changes in one of the main components of the picture, such changes being detected more often and quicker. Even then, it takes 7 seconds on average to detect central changes, and once participants become aware of them, it becomes striking and almost impossible to avoid.

The neural mechanisms at the root of error detection and expectation violation are studied using electroencephalogram (EEG) event related potential (ERP) components called error related negativity (ERN) and mismatch negativity (MMN). ERN usually occurs when someone commits an error and it is a reaction attributed to the anterior cingulate cortex (Holroyd and Coles, 2002). Its detection can be due either to the presentation of the stimuli or to the response and it is usually detectable even when the individual is unaware, they made the mistake. MMN on the other hand, usually occurs between 250 and 450 milliseconds after conflicting stimuli are presented, and is seen as an automatic reaction that triggers a redirection of attention towards the detected conflict (Garrido et al., 2009). It has been shown to be defective in pathologies where false conclusions are made such as schizophrenia (Umbricht and Krljesb, 2005), or dyslexia (Bishop, 2007). Furthermore, larger MMN in typical individuals lead to better and faster error detection (Garrido et al., 2009), pointing towards a causal role of this ERP.

In the military domain, where dynamic decision-making is necessary, Vachon and colleagues (2012) studied change blindness in a computerized simulation where participants needed to make tactical decisions on a naval warfare ship with many aircrafts surrounding it. They documented implicit change blindness with eye-tracking. Although such changes do not specifically need to be reported as detected in the task, they can lead to detrimental behavior if they are not properly addressed. There are many similarities between such behavioral patterns and the tasks of a CFI, in that they need to monitor and detect a variety of changes with different levels of importance. Dehais and colleagues (2014) on the other hand, observed change deafness (the auditory variant of change blindness) in pilots in a simulated cockpit, and there are no reasons why such phenomena could not also be observed in CFIs.

Risk tolerance

Once an error or a deviation from expectations is detected by the CFI, it has to be adequately acted upon to correct the student's behavior. This capacity to engage in corrective actions is dependent on risk tolerance. Hunter (2002) defined risk tolerance as the amount of risk an individual is willing to take in order to achieve a goal. He also found that risk tolerance can significantly impact a pilot's decision-making. Ji and colleagues (2011) demonstrated that high scores on risk tolerance were associated with both a hazardous attitude and an adoption of less safe behavior that may lead to less safe flight operation. Pilots who had high scores in risk tolerance measures, tended to perceive the risk related to aviation accidents as low, besides having a positive hazardous attitude (Ji et al., 2011). Therefore, detecting the degree of a pilot's risk tolerance is essential for determining their behavior and future action during a flight. While aviation regulation agencies ensure that both the pilot and the CFI are aware of and fully comprehend flight safety standards, tolerance to deviations in these standards could have detrimental effects on the training and evaluation of pilots.

Around 70% to 80% of flight accidents are caused by human errors (Wiegmann and Shappell, 2017). Therefore, it is crucial for flight instructors to take this factor into consideration. CFIs must ensure that the pilot doesn't make a maneuver that could have a negative impact on the flight. Detection of risky pilot behavior is critical to the improvement of flight safety. An instructor who fails to detect or acknowledge risky behavior of a pilot may endanger the safety of the flight.

Risk tolerance leads to different interpretations of human error (Baysari et al., 2009). Measuring the flight instructor's risk tolerance level might help us understand how he interprets a pilot's risky behavior or maneuver during a simulation. During a flight simulation session, the CFI's actions should be directed towards managing any negative behavior from the pilot that could contribute to aviation accidents. A CFI who does not either notice or respond with sufficient strictness to risky behavior is less efficient than one who does and may endanger the safety of the flight.

Risk tolerance is an attitude that can be measured both explicitly and implicitly. Explicit attitudes towards a person or a subject can be assessed using a classic questionnaire, but the fact that these attitudes are not immune to social desirability biases enhances the interest for methods that could circumvent such issues. Implicit biases, on the other hand, are unconscious attitudes and associations that can affect behavior without being aware of the causal impact of this bias in decision-making.

There is a stronger correlation between different measures of explicit attitudes than between different measures of explicit and implicit attitudes, which goes to show that they do bring a new set of evidence to assess driving causes of behaviors (Hahn et al., 2014). These biases have been shown to impact the work of health professionals (Green et al., 2007), members of law enforcement (Kang, Bennett and Carbado, 2012) and teachers (van den Bergh et al., 2010), among other professions. There is no reason to believe that CFIs might be immune to these biases, thereby impacting the interactions they have with pilots.

The implicit-association test (IAT) is a widely-used experimental paradigm to assess implicit attitudes (Greenwald, McGhee and Schwartz, 1998; Greenwald et al., 2009). It is based on the principle that if a person implicitly associates two concepts together, the behavioral output will be facilitated when compared to information that is conflicting to the direction of the bias. This is usually implemented using an image or a word at the center of a screen that has to be categorized in either a descriptive way (black or white) or a judgmental way (pleasant or unpleasant) with one of the words from each category either to the left or the right of the screen, indicating the required response keys. By changing the descriptive and the judgmental words that are presented in the same corner and contrasting reaction times between conditions, behavioral facilitation can be assessed, and a bias score can be extracted.

In the aviation domain, a study by Molesworth and Chang (2009) showed that a risk-taking variation of the IAT was the strongest predictor of risk-taking behavior in pilots, when compared to risk perception and risk behavior questionnaires, and also when compared to the pilot's experience, which were concepts previously mentioned as impacting risk-taking behavior. If such a link exists in pilots, it may also exist in trainers, where an implicit attitude skewed towards a high risk-taking tendency might lead to the tolerance of more risky behavior from the trainee. Yet, risk-taking is not the only aspect of aviation where implicit attitudes might have an impact on the instructor's performance. The attitude towards different characteristics from students could certainly be probed using this technique, given the large body of literature on the subject, and characteristics of the technological systems such as trust in its automation have also led to interesting insights in the automation of routine airport security procedures (Merritt et al., 2013). This paradigm has recently been used in combination with eye-tracking (Mele and Federici, 2012; Mele, Federici and Dennis, 2014), revealing longer fixations in situations that contradicted the implicit bias of participants.

Flow and Cognitive Absorption

Another concept that has recently been linked to situation awareness and that has received a substantial amount of interest in the scientific literature is flow (Shneiderman and Benderson, 2005). The state of flow was first conceptualized by Csikszentmihalyi (1975/2000) as an optimal human experience. It is a subjective experience that consists of equilibrium between a higher than average perceived skillset and a task that is perceived as challenging. The relationship between flow and task performance is both direct and indirect. Two components of flow, concentration and control, can enable increases in performance (Eklund, 1996) and since flow is a state which is sought through intrinsic motivation, it indirectly influences performance by increasing motivation and engagement towards a task (Engeser and Rheinberg, 2008).

The relation between flow and human computer interactions (HCI) was first traced by Finneran and Zhang (2003, 2005) and by Agarwal and Karahanna (2000) and it was coined cognitive absorption (CA). CA is considered as an antecedent that influences technology acceptance (Saadée and Bahli, 2005). CA variables are described as a form of internal motivation, given that interacting with IT may lead to pleasure and satisfaction (Vallerand, 1997). The state of involvement with IT is a variable that led to perceived ease of use and perceived usefulness. In the context of an e-learning course, Saadé and Bahli (2005) used a questionnaire, based on a five-point Lickert-type scale, to measure perceived usefulness, perceived ease of use and their relationship to CA. They concluded that "cognitive absorption was shown to be an important antecedent to perceived usefulness but less important to perceived ease of use" (Saadée and Bahli, 2005). Therefore, the degree of involvement with IT is considered to play an important role in determining the perception of use and may lead to a better perception of easiness a user gains from using IT.

The two main scales developed and used to measure CA are the one developed by the original authors (Agarwal and Karahanna, 2000) and a scale developed by Barki, Paré and Sicotte

(2008). Flow has traditionally been measured via questionnaires like the flow questionnaire, the experience sampling method and the standardized scales of the componential approach (Moneta, 2012).

Recent developments by Léger and colleagues (2014) were the first to identify physiological correlates of CA in a training context. CA correlates positively with the EEG alpha-frequency band, which is linked to a relaxed state, and negatively with the beta-frequency band, which is linked to vigilance. It is also correlated with lower EDR variability and heart rate. Biometric tools could measure flow state in other contexts than in interaction with technology. Electrocardiograph (ECG) electrodes or galvanic skin response (GSR) can measure the physiological state of a participant, while interbeat interval (IBI) standard deviation and galvanic skin response (GSR) can measure participant challenge states from boredom to flow (Afergan, 2015). Evidence of a positive relation between stress-hormone cortisol and flow experience was recently demonstrated using ECG (Peifer et al., 2014) and PET scans were used to find a positive correlation between dopamine production and flow proneness (de Manzano et al., 2013). Cardiac and respiratory activity measures were used to partially demonstrate that flow peak activity is associated with the relaxation state, the parasympathetic state (Harmat et al., 2015).

In the gaming industry, scenarios that adjust difficulty levels according to the performance of the participant tend to keep the player more immersed, than those that simply adjust difficulty levels over the course of gameplay (Engeser and Rheinberg, 2008; Kennedy et al., 2010). Following these results, it might be crucial for instructors to adapt flight simulation sessions in order to maximize the student's flow experience.

Workload

During a high workload period, the need to perform complex tasks and multiple tasks simultaneously can diminish the performance of one or all tasks (Loukopoulos, Dismukes and Barshi, 2009). Hicks and colleagues (2014) mentioned that "if one or both pilots experience excessively high workload while performing flight and mission tasks, the tasks may be performed ineffectively or abandoned." In the aviation domain, pilots and flight instructors monitor a vast number of features from the aircraft system and face scenarios (cross-wind landing, motor failure, etc.) that may exceed their workload limitation. Bearing this in mind, CFIs should perhaps evaluate the workload degree of their pilots and adapt the flight simulation training to help the latter manage their workload level and deal with potential overload to maximize the learning process of the student.

During periods of low workload, the instructor should also take into consideration the fact that pilots may become less vigilant about their flight duties. This can also apply to CFIs, as low workload periods may lead to boredom and lack of vigilance (Boyer et al., 2015) which, in turn, results in a lack of performance on the flight instructor's part. Given the potential impact of workload on performance, the goal is to conduct a training simulation that allows the pilot and the CFI to remain in a balanced zone of cognitive load.

Cognitive load theory, introduced in the 1980s, can be defined as "the required working memory resources needed to process the information with the resources dealing with the division of cognitive load" (Sweller, Ayres and Kalyuga, 2011b). Cognitive load can be divided into intrinsic cognitive load, extraneous cognitive load and germane load (Sweller, Ayres and Kalyuga, 2011b). Intrinsic cognitive load implies that each instruction has a level of difficulty by itself. Extraneous cognitive load is managed by instructional designers and refers to the way that an information is assimilated or presented (an instructor can either verbally explain or visually demonstrate a maneuver). Germane cognitive load refers to the processing, construction and automation of schemas (Sweller, Ayres and Kalyuga, 2011b). People exceed their cognitive load limitation if the resource that they are dealing with exceeds their working memory load capacity. The definition of cognitive workload and mental workload are similar and they both refer to the ability to deal with simple or complex tasks.

Workload measurement is essential for establishing "the mental cost of performing a task to predict operator or system performance" (Cain, 2007). Workload measurement can include self-response questionnaires, performance measures and biometric tools. For simulation-based training, Farmer and Brownson (2003) recommended that a series of workload measures should be used, such as: Modified Cooper-Harper (MCH); Instantaneous Self-Assessment (ISA); Primary and Secondary tasks; Heart Rate; Heart Rate Variability; NASA-TLX; Defence Research Agency Workload Scale (DRAWS); and Blink rate. Brad Cain did a review of the mental workload literature (2007) and found that the most popular method of measuring workload was to use a global and univariate workload measure with NASA-TLX or Subjective Workload Assessment Technique (SWAT) and a set of secondary task measures.

The Modified Cooper-Harper (MCH) scale was introduced in the aviation domain. It is a uni-dimensional measure that requires a decision tree to measure operator mental workload and it is based on the theory that "there is a direct relationship between the difficulty of aircraft controllability and pilot workload". The Instantaneous Self-Assessment (ISA) scale is a tool that requires participants to undertake a subjective workload evaluation during a primary task such as air traffic control. The DRAWS is a subjective workload assessment scale (De Maio and Hart, 1999).

Stress

Stress has been found to have a negative impact on flying skills involving psychomotor, working memory, and attentional components (Stokes and Kite, 1994). CFIs can decrease the negative impact of stress by using simulator-based stress training which can improve pilot performance (Mcclernon, 2009; McClernon et al., 2011). Similar to the concept of workload, CFIs must not only manage the stress level of the pilots to optimize their performance, but also maintain their own stress level in a balanced zone to prevent boredom from setting in during the training simulation.

Selye (1956) was the first person to introduce the concept of stress, which he defined as "a nonspecific result to any demand upon the body". Stress is a process engendered by environmental demands in which the perceived demand exceeds the resources of the individual facing this situation. The process causes an undesirable psychological, physiological, or behavioral outcome to the person (Saunders et al., 1996).

The relationship between stress and performance was first reviewed by Gmelch and colleagues (1982). Their work demonstrates an inverted U relationship between performance and stress. A low degree of stress causes boredom, frustration, fatigue and dissatisfaction, resulting in a negative effect on performance. A high degree of stress causes irrational problem solving, exhaustion, illness and low self-esteem, resulting in a negative effect on performance. To optimize performance, a person must be in an optimum simulation zone, implying that the level of stress is neither too low nor too high (Gmelch et al., 1982).

There are currently many ways to measure a person's stress level. There are four types of stress indices, namely subjective, behavioral, psychophysiological and biochemical (Stokes and Kite, 1994). Subjective measures require a questionnaire, in which a person writes down how he feels and believes he is doing. Behavioral measures may require a range of computerized test batteries, specialized performance tests and flight simulator training. Psychophysiological indices can be measured via the heart rate, skin conductance, respiratory rate, muscle tension, etc. Biochemical indices are objective measures of neurotransmitters and their metabolites, like serotonin, epinephrine, norepinephrine and dopamine. Electrodermal activity (EDA) measures the skin conductance and has been proven to effectively measure the stress level of a person.

Discussion about the scoping review

This review gives an overview of the concepts related to the evaluation of CFI performance, which has theoretical and practical implications. The practical implications of this review are relevant for the aviation industry. It provides a clear understanding of CFI behavior factors that need to be optimized in order to improve the skills of the CFI, which could lead to the development of better simulation training courses and the development of better pilots. We also provided substantial evidence that very few studies have directly focused on CFI behaviors, and particularly those related to situation awareness. This review also offers state of the art measures that could be used for each concept. Researchers may also benefit from our review as it contains a list of validated measures that could be used in future experiments to evaluate CFI performance.

In this review, we investigated the concepts that could improve our knowledge of CFI performance factors, with special emphasis on situation awareness and related constructs. We consider that the selection of the concepts and constructs presented in this review represent one of the limits of this study. To our knowledge, there is no consensus on establishing which are the specific performance factors of a CFI. The aim of the article was to do an exhaustive review of the theme present in the literature based on the systematized framework of Thomas and colleagues (2000). Other factors related to the human factors and eras decision-making, fatigue or mental models could have been applied to this review. Therefore, our review provides a global exploratory list of the concepts that could be applied to evaluate a CFI's performance.

Conclusion

Only 20 peer-reviewed articles identified performance factors of CFIs, which shows that this field of research is still in its early stages, even though the oldest articles date back to 1965. Validated psychometric and biometric tools could provide a clear understanding of the attentional and decisional processes at play while developing situation awareness in a dynamic environment and help quantify the task load and factors affecting it. Connecting the dots between the possible performance factors developed in this review will enable the aviation research community to develop and inform best practices in the selection, training and expertise development of CFIs. A better understanding of CFI performance factors in a training simulation context might help to increase the quality of the CFI’s teaching, which might lead to an upsurge in the quality of the pilot’s learning process and the development of better pilots.

References

Afergan, D. A. (2015). *Implicit Brain-Computer Interfaces for Adaptive Systems: Improving Performance through Physiological Sensing* (Doctoral dissertation, Tufts University).

Agarwal, R. and Karahanna, E. (2000). Times flies when your're having fun: cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24(2), pp. 665–694.

Aghaei Chadegani, A. *et al.* (2013). A comparison between two main academic literature collections: Web of science and scopus databases. *Asian Social Science*, 9(5), pp. 18–26.

Arksey, H. and O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*, 8(1), pp. 19–32.

Ball, K. and Owsley, C. (1993). The useful field of view test: a new technique for evaluating age-related declines in visual function. *Journal of the American Optometric Association*, 64(1), pp. 71–79.

Barki, H., Paré, G. and Sicotte, C. (2008). Linking IT implementation and acceptance via the construct of psychological ownership of information technology. *Journal of Information Technology*, 23(4), pp. 269–280.

Baysari, M. T., Caponecchia, C., McIntosh, A. S., & Wilson, J. R. (2009). Classification of errors contributing to rail incidents and accidents: A comparison of two human error identification techniques. Safety Science, 47(7), 948-957.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*(2), 497-527.

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: where are we, and where should we be going?. *Psychological Bulletin*, 133(4), pp. 651–672.

Bjerke, E. and Malott, D. (2011). Impacts of Public Law 111-216: Will the flight instructor career path remain a viable option for aspiring Airline pilots?. *Collegiate Aviation*

Review, 29(1), pp. 1–9.

Boyer, M., Cummings, M. L., Spence, L. B., & Solovey, E. T. (2015). Investigating mental workload changes in a long duration supervisory control task. Interacting with Computers, 27(5), 512-520.

Burian, B. K. and Feldman, J. (2009). Certified flight instructor weather training: Perspectives and practices. *International Journal of Aviation Psychology*, 19(3), pp. 217–234.

Byrnes, K. P. (2017). Employing Flight Simulation in the Classroom to Improve the Understanding of the Fundamentals of Instruction among Flight Instructor Applicants. *Journal of Aviation/Aerospace Education & Research*, *26*(1), 49.

CAE (2017) *Airline Pilot Demand Outlook Airline Pilot Demand: 10-year view*.

Cain, B. (2007). A Review of the Mental Workload Literature. *Defence Research and Development Toronto (Canada)*, (1998), pp. 4-1-4–34.

Corl, L. and LaRoche, J. (2001) *Predicting Human Performance*. Saint-Laurent, Quebec: Helio Press.

Crow, B., Niemczyk, M., Andrews, D., & Fitzgerald, P. (2011). Role playing in flight instructor training: How effective is it?. International Journal of Applied Aviation Studies, 11(1), 1-12.

Csikszentmihalyi, M. (2000) *Beyond Boredom and Anxiety*. San Francisco: (Original work published in 1975).

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. Human factors, 56(4), 631-644.

Diels, E. A. (2007) *Moral Development in Pilot Populations*.

Eklund, R. C. (1996). Preparing to Compete: A Season-Long Investigation with Collegiate Wrestlers. *The Sport Psychologist*, 10(2), pp. 111–131.

Ellis, G. A., & Roscoe, A. H. (1982). *The Airline Pilot's View of Flight Deck Workload: A Preliminary Study Using a Questionnaire* (No. RAE-TM-FS (B)-465). ROYAL AIRCRAFT ESTABLISHMENT FARNBOROUGH (ENGLAND).

Endsley, M. R. (2015). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1), pp. 4–32.

Engeser, S. and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3), pp. 158–172.

Farmer, E. and Brownson, A. (2003). Review of workload measurement, analysis and interpretation methods. *European Organisation for the Safety of Air Navigation*, 47(25), p. 33.

Finneran, C. M. and Zhang, P. (2005). Flow in Computer-Mediated Environments: Promises and Challenges. *Communications of the Association for Information Systems*, 15, pp. 82–101.

Finneran, C. and Zhang, P. (2003). A Person-Artifact-Task (PAT) Model of Flow Antecedents in Computer- Mediated Environments. *International Journal of Human-Computer Studies*, pp. 1–30.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. Clinical neurophysiology, 120(3), 453-463.

Gibb, R., Gray, R. and Scharff, L. (2010) *Aviation visual perception: Research, Misperception and Mishaps*. Ashgate Publishing Ltd.

Gibbs, R., Davies, G. and Chou, S. (2016). A systematic review on factors affecting the likelihood of change blindness. *Crime Psychology Review*. Routledge, 0(0), pp. 1–20.

Gmelch, W. H. (1982). What Stresses School Administrators--And How They Cope.

Gontar, P. and Hoermann, H.-J. (2015). Interrater Reliability at the Top End: Measures of Pilots' Nontechnical Performance. *The International Journal of Aviation Psychology*. Taylor & Francis, 25(3–4), pp. 171–190.

González Vega, N. (2002) *Factors Affecting Simulator-training Effectiveness*.

Gray, R., Gaska, J. and Winterbottom, M. (2016). Relationship between sustained, orientated, divided, and selective attention and simulated aviation performance: Training & pressure effects. *Journal of Applied Research in Memory and Cognition*. The Society for Applied Research in Memory and Cognition, 5(1), pp. 34–42.

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., &

Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. Journal of general internal medicine, 22(9), 1231-1238.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of personality and social psychology, 97(1), 17.

Greenwald, A. G., McGhee, D. E. and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), pp. 1464–1480.

Guion, R. M. and Ikomi P. A. (2000). The Prediction of Judgment in Realistic Tasks: An Investigation of Self-Insight. *The International Journal of Aviation Psychology*, 10(2), pp. 135–153.

Guz, A. N. and Rushchitsky, J. J. (2009). Scopus: A system for the evaluation of scientific journals. *International Applied Mechanics*, 45(4), pp. 351–362.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. Journal of Experimental Psychology: General, 143(3), 1369.

Hale, S. K. and Breaux, R. (2011). Enhancing Pilot Training with Advanced Measurement Techniques. in *9th International Conference, EPCE 2011 Held as Part of HCI International 2011 Orlando, FL, USA, July 9-14, 2011 Proceedings*, pp. 540–545.

Hall, J. C. (2011). Survey of flight instructors' experiences in communication training. *Collegiate Aviation Review*, 29(1), pp. 45–56.

Harbeck, T., Kirschner, J., Wulle, B., & Bowen, E. (2014). Evaluating Flight Instructor Perceptions of Light Sport Aircraft. Collegiate Aviation Review, 32(1), 33.

Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., & Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. International Journal of Psychophysiology, 97(1), 1-7.

Hicks, J. S., Durbin, D. B., Morris, A. W., & Davis, B. M. (2014). A Summary of Crew Workload and Situational Awareness Ratings for US Army Aviation Aircraft (No. ARL-TR-6955). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE.

Holroyd, C. B. and Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), pp. 679–709.

Hoover, A. L. (2008). Educational Learning Theories: Informing the Fundamentals of Instruction. *International Journal of Applied Aviation Studies*, 8(2), pp. 363–371.

Hunter, D. R. (2002) *Risk Perception and Risk Tolerance in Aircraft Pilots*.

Ji, M., You, X., Lan, J., & Yang, S. (2011). The impact of risk tolerance, risk perception and hazardous attitude on safety operation among airline pilots in China. Safety science, 49(10), 1412-1420.

Kang, J., Bennett, M., Carbado, D., & Casey, P. (2011). Implicit bias in the courtroom. UCLa L. rev., 59, 1124.

Kennedy, Q., Taylor, J. L., Reade, G., & Yesavage, J. A. (2010). Age and expertise effects in aviation decision making and flight control in a flight simulator. Aviation, space, and environmental medicine, 81(5), 489-497.

King, W. R. and He, J. (2005). Understanding the Role and Methods of Meta-Analysis in IS Research. *Communications of the Association of Information Systems*, 16(October), p. 654.

Kreienkamp, R. A. and Luessenheide, H. D. (1985). Similarity of Personalities of Flight Instructors. *Psychological Reports*, 57(2), pp. 465–466.

Léger, P. M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. Computers in Human Behavior, 34, 273-283.

Lépinard, P. (2014). Du serious gaming au full flight simulator: proposition d'un cadre conceptuel commun pour la formation des formateurs en simulation. Systèmes d'information & management, 19(3), 39-68.

Loukopoulos, L. D., Dismukes, K. and Barshi, I. (2009). The Multitasking Myth: Handling Complexity in Real-World Operations. *International Journal of Applied Aviation Studies*, 9(1), pp. 109–116.

De Maio, J. and Hart, G. S. (1999) *Awareness and Workload Measures for SAFOR*.

Mann, D. T., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. Journal of Sport and Exercise Psychology, 29(4), 457-478.

de Manzano, Ö., Cervenka, S., Jucaite, A., Hellenäs, O., Farde, L., & Ullén, F. (2013). Individual differences in the proneness to have flow experiences are linked to dopamine D2-receptor availability in the dorsal striatum. Neuroimage, 67, 1-6.

McClernon, C. K. (2009). Stress effects on transfer from virtual environment flight training to stressful flight environments. NAVAL POSTGRADUATE SCHOOL MONTEREY CA.

McClernon, C. K., McCauley, M. E., O'Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. Human factors, 53(3), 207-218.

McDale, S. and Ma, J. (2008). Effects of Fatigue on Flight Training: A Survey of U.S. Part 141 Flight Schools. *International Journal of Applied Aviation Studies*, 8(2), pp. 311–337.

Mele, M. L. and Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive Processing*, 13(1 SUPPL).

Mele, M. L., Federici, S. and Dennis, J. L. (2014). Believing is seeing: fixation duration predicts implicit negative attitudes. *PloS one*, 9(8), p. e105106.

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. Human factors, 55(3), 520-534.

Molesworth, B. R. C. and Chang, B. (2009). Predicting Pilots' Risk-Taking Behavior Through an Implicit Association Test. *Human Factors*, 51(6), pp. 845–857.

Moneta, G. B. (2012). On the Measurement and Conceptualization of Flow. *Advances in Flow Research*, pp. 23–50.

Nählinder, S. and Berggren, P. (2005). Increasing Training Efficiency Using Embedded. *International Journal*, pp. 2197–2200.

O'Brien, K. S. and O'Hare, D. (2007). Situational awareness ability and cognitive skills training in a complex real-world task. *Ergonomics*, 50(7), pp. 1064–1091.

O'Donnell, R. D., Moise, S., & Schmidt, R. M. (2005). Generating performance test batteries relevant to specific operational tasks. Aviation, space, and environmental medicine, 76(7), C24-C30.

Van Orden, K. F., Jung, T. P. and Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, 52(3), pp. 221–240.

Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. Information & Management, 52(2), 183-199.

Peifer, C., Schulz, A., Schächinger, H., Baumann, N., & Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress—can u shape it?. Journal of Experimental Social Psychology, 53, 62-69.

POMAROLLI, R. S., & Ambler, R. (1965). Voluntary withdrawal from primary flight training as a function of the individual flight instructor (No. SR-65-2). NAVAL SCHOOL OF AVIATION MEDICINE PENSACOLA FLA.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), pp. 3–25.

Rensink, R. A., Regan, J. K. O. and Clark, J. J. (1997). To See or Not to See : The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, 8(5), pp. 368–373.

Saadé, R. and Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: An extension of the technology acceptance model. *Information and Management*, 42(2), pp. 317–327.

Saunders, T., Driskell, J. E., Johnston, J. H., & Salas, E. (1996). The effect of stress inoculation training on anxiety and performance. Journal of occupational health psychology, 1(2), 170.

Selye, H. (1956) *The Stress of Life*.

Shneiderman, B. and Benderson, B. B. (2005). Maintaining concentration to achieve task completion. *DUX '05 Proceedings of the 2005 conference on Designing for User eXperience*, pp. 2–7.

Sohn, S. Y. and Jo, Y. K. (2003). A study on the student pilot's mental workload due to personality types of both instructor and student. *Ergonomics*, 46(15), pp. 1566–1577.

Stokes, A. and Kite, K. (1994) *Flight Stress: Stress, Fatigue and Performance in Aviation*. 1st edn. Gower Technical.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In Cognitive load theory (pp. 71-85). Springer, New York, NY.

Talleur, D. A. and Wickens, C. D. (2003). The Effect of Pilot Visual Scanning Strategies on Traffic Detection Accuracy and Aircraft Control. in *12th International Symposium on Aviation Psychology*. Dayton OH, p. 6.

Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational Awareness (pp. 111-128). Routledge.

The Boeing Company (2017). *2017 Pilot & Technician Outlook Pilot & Technician Outlook*.

Thomas, M. J. W. (2000). Enhancing Instructional Systems : The Development of a Tool for Evaluating Instructor and Student. *Transport*, pp. 41–51.

Umbricht, D. and Krljesb, S. (2005). Mismatch negativity in schizophrenia: A meta-analysis. *Schizophrenia Research*, 76(1), pp. 1–23.

Vachon, F., Vallières, B. R., Jones, D. M., & Tremblay, S. (2012). Nonexplicit change detection in complex dynamic settings: What eye movements reveal. Human Factors, 54(6), 996-1007.

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In *Advances in experimental social psychology* (Vol. 29, pp. 271-360). Academic Press.

Wetmore, M. and Lu, C. (2007). The Effects of Pedagogical Paradigms on Aviation Students with Hazardous Attitudes. *Journal of Aviation/Aerospace Education & Research*, 16(3), pp. 25–36.

Wickens, C. D., Gempler, K. and Ephimia, M. M. (2000). Workload and Reliability of Predictor Displays in Aircraft Traffic Avoidanc. *Transportation Human Factors*, 2(2), pp. 99–

126.

Wiegmann, D. A. and Shappell, S. A. (2017). *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*. Routledge.

Xu, J., Sohoni, M., McCleery, M., & Bailey, T. G. (2006). A dynamic neighborhood based tabu search algorithm for real-world flight instructor scheduling problems. European Journal of Operational Research, 169(3), 978-993.

Chapitre 3 : Article 2

Extending the cognitive fit theory to analyse the effect of simultaneously available dynamic information representations on certified flight instructor (CFI) performances: the development of an experimental paradigm

Christophe Lazure[1], Laurence Dumont[1], Sylvain Sénécal[1], Jean-François Delisle[2], Shang Lin Chen[1] et Pierre-Majorique Léger[1]

[1]HEC Montréal, Montréal, Canada

{christophe.lazure,laurence.dumont,shang-lin.chen,pml,ss}@hec.ca

[2]Polytechnique de Montréal, Montréal, Canada

Jean-francois.delisle@cae.com

Abstract

A. Objective: Based on cognitive fit theory, we aim to develop an experimental paradigm to analyze the effect of information representations on certified flight instructor (CFI) performance when multiple types of dynamic information representations are simultaneously available to the CFI. We conducted a pilot study to demonstrate the potential and the feasibility of this paradigm.

B. Background: Prior research on the cognitive fit theory used the cognitive fit framework for sequential analyses, where only one information representation type is evaluated at once. When considering the usual interfaces available to CFI, different dynamic representations are presented simultaneously. The cognitive fit theory needs to be extending to analyses the effect of simultaneously presented dynamic information representations on CFI performance.

C. Method: The proposed paradigm was tested in a laboratory experiment in which a within-subjects experimental design was used to evaluate a series of landing and takeoff manoeuvres.

D. Result: Different visual scan patterns through the information representations affected the speed at which CFI conduct their evaluations of manoeuvre and their cognitive load.

E. Conclusion: The results of our experiment provide initial support the experimental paradigm and call for testing it in subsequent studies to confirm the potential of this approach.

Keywords: Certified flight instructor; cognitive fit theory, information representation; visual attention; cognitive load; visual scan pattern;

Introduction

In the field of aviation, training sessions are generally executed in a simulation training context. In these settings, certified flight instructors (CFIs) are responsible for evaluating the student pilot's performance. To do so, they use different interfaces and sources of information to determine if the student pilot's actions are safe and meet the standards of the aviation industry. State of the art interfaces uses a combination of spatial representations (e.g. two-dimensional, three-dimensional representations and flight control) and symbolic representations (flight instruments and information table) to present the data of the flying lesson.

While aviation research has always had a strong rooting in HCI, most of the past literature published in this field has been focused on flight pilots (e.g., Alexander et al., 2005; Wickens, 2002). Only a small amount of peer-reviewed papers – 20 to be exact – have been specifically focused on the CFIs (Lazure et al., 2018). In order to meet the industry's demand for pilots and to diminish the lack of research on antecedents of CFI performance such as CFI information processing (Peißl, Wickens, & Baruah, 2018; Vidulich et al., 2010), we argue that research investigating how CFIs process information presented in different formats (specifically spatial and symbolic formats) is a worthwhile topic to investigate at this point in time.

The relative superiority of an information representation over another can be understood through the cognitive fit theory. To our knowledge, this framework has not yet been applied in the aviation domain (Bacic, 2014). The cognitive fit theory states that information representations must support the problem-solving processes in order to reduce problem solving effort. Vessey (1991) founding study of the theory, the importance of considering the interaction between the representation of a problem and the task at hand is highlighted. The emphasis on this interaction facilitates the understanding of information, thus improving decision abilities. The task type must match the problem representation type for efficient and effective task completion.

Prior research on the cognitive fit theory used the theory for sequential analyses, where only a single information representation type is evaluated at once (Bacic, 2014; Vessey, 2006). When considering the usual interfaces available to CFIs, different dynamic representations are presented simultaneously. Thus, it is necessary to evaluate if the cognitive fit theory can be applied when multiple information representations are simultaneously available to the CFI. Therefore, this article answers the following research question: In a context where dynamic information representations are simultaneously presented to users, are their visual patterns associated with better performance, thus providing evidence of cognitive fit?

Based on the cognitive fit theory, we aim to develop an experimental paradigm that investigates how the dynamic use (visual patterns) of the different information representations available to the CFI affect his/her performance during the evaluation of a student pilot's manoeuvre. To demonstrate the potential and the feasibility of this paradigm, we conducted a correlational study. This laboratory experiment employed a combination of self-reported (explicit) measures and psychophysiological (implicit) measures.

Explicit measures (e.g., self-reported questionnaires) can produce an incomplete picture of the user's state when considered alone, because they provide a subjective evaluation and are administered after a task, which makes them prone to many measurement biases. To evaluate the subject's attentional and cognitive processes in real time, implicit measures were also be used. The use of biometric tools allow the capture of signals from the user's reactions. When used together, explicit and implicit measures produce more valid and reliable conclusions than when either measurement type is used alone (Ortiz de Guinea et al., 2014).

This paper provides exploratory results on which type of information representation, in a complex evaluation task requiring the overall assessment of the student's performance, leads to a better cognitive fit in term of decision-making. This study contributes to the literature by extending the cognitive fit theory by analysing the effects of information representations on performance, when multiple information representations are simultaneously available to the user. This paper provides insight to practitioners working in the field of aviation on how to design interfaces that better fit the tasks of CFIs.

Literature Review

Cognitive Fit Theory

Cognitive fit theory was first introduced in response to the inconsistent results across prior studies which do not clearly demonstrate if information processing was more efficiently using graphical representation or tables, thus demonstrate which representation provides a systematic advantage of a type of representation over the other (Vessey, 1991). Cognitive fit theory resolves this issue by suggesting that it is not the intrinsic nature of a representation that provides an advantage, but the interaction between a problem representation and a task that determines problem-solving efficiency. According to the founding study of the theory (Vessey, 1991), there is a relationship between presentation format and task type. Problem representation must match a user's decision-making strategy in the realization of a task (John & Kundisch, 2015; Vessey & Galletta, 1991). Cognitive fit occurs between the presentation format and task type "when the problem-solving aids support the task strategies required to perform that task" (Vessey, 1991, p. 220). When a good cognitive fit between the problem representation and the task is achieved, both accuracy and completion time are improved (Khatri et al., 2006).

To classify the relationship between presentation format and task type, Vessey (1991) suggested considering the cognitive types as two categories: spatial and symbolic. Symbolic presentation format emphasizes information on discrete data values while spatial presentation format emphasizes the problem area as a whole (Vessey & Galletta, 1991). Symbolic task involves extracting discrete and precise data values (e.g extracting the speed of the aircraft at a precise moment) while spatial tasks could require making associations or identifying perceiving relationships in the data (e.g. evaluate how the fuel consumption of the aircraft is affected by its speed) (Vessey & Galletta, 1991). Spatial tasks are best supported by spatial representations, and symbolic tasks by symbolic representations. As mentioned by Vessey and Galletta (1991, p. 67) "cognitive fit results when the problem representation and the task both emphasize the same type of information."

Visual Attention

Prior research has evaluated the cognitive fit theory trough sequential analyses, where only one of the two cognitive types (spatial or symbolic) are evaluated at a time (Bacic, 2014; Vessey, 2006). On a CFI interface, the two cognitive types are presented simultaneously. In a context where many representations of the information are available, to make a decision, it is necessary to be able to identify which representation is used. We argue that if the relative use of a representation over another is associated with a better performance, one can conclude that the cognitive fit was better. Such measurement of information representation usage can be conducted through an analysis of the visual attention.

Visual attention refers to the capacity to focus our attention on a specific object in the visual environment (Carrasco, 2011). Visual attention guides eye movements in the search for relevant information in the environment, plays an important role in the problem-solving process and guides individual's behaviour (Carrasco, 2011). Guided by visual attention, eye movements are essential to the selection and processing of different targets (Carrasco, 2011). To make a decision based on multiple targets in the environment, both saccades and smooth pursuits are required (Carrasco, 2011). Saccadic eye movements are used to move attention from one target to another, while keeping track of the mobile targets by smoothly following the movements of the target (Kowler, 2011). Measuring saccadic movements captures the fixation counts and the duration of a subject's in a specific area (Holmqvist et al., 2011). Eye tracking technologies can be used to measure these eye movements (Riedl & Léger, 2016).

Eye-tracking measures have been used in the field of aviation to improve the human-computer interaction and the flight safety (Dehais et al., 2008; Peißl et al., 2018; Ziv, 2016). In a flight simulation training context, Yu and colleagues (2014), have suggested that eye-tracking technologies provide real-time information about the pilot's visual scan pattern and can thus improve training effectiveness. Past literature has mainly focused on the flight pilot, yet we argue that CFI's visual scan patterns should also be considered in order to improve the training session effectiveness.

Cognitive Load

Past research on the cognitive fit theory has used cognitive effort as a foundation construct for the theory (Bacic, 2014; Vessey, 2006). Cognitive load refers to the cognitive effort used in a task, and can be defined as "the required working memory resources needed to process the information with some resources dealing with intrinsic cognitive load (germane resources) and other resources dealing with extraneous cognitive load (extraneous resources)" (Sweller, Ayres, & Kalyuga, 2011, p. 58). Hence, there are three main types of cognitive load: (1) intrinsic cognitive load, in which working memory is solicited by the intrinsic nature of the information; (2) extraneous cognitive load, in which working memory is solicited by the nature of the information transmission and how it is presented to the user; and (3) germane cognitive load, which was introduced to explain the necessity of cognitive effort in learning activities (Sweller et al., 2011). Since germane cognitive load is imposed during learning activities and that CFIs are experts in their domain, germane cognitive load was not considered in the context of our research.

For example, a CFI is receiving information about a difficult flight manoeuvre and that he must evaluate the performance of the manoeuvre. A flight manoeuvre can require high levels of working memory to be evaluated, which is based on its intrinsic complexity, and will therefore elicit a high intrinsic cognitive load. Now, let's assume that the same manoeuvre is presented to the CFI through two types of information representations: a 3D video of the flight manoeuvre (spatial representation) and a flight instrument (symbolic representation). The evaluation of the flight manoeuvre will require different levels of working memory based on how the information is presented. Hence, extraneous cognitive load is impacted by the type of information representation.

The cognitive fit theory states that cognitive effort should be lower when a problem representation matches an associated task (Vessey, 1991). However, few papers have assessed the effect of cognitive load on performance (Bacic, 2014). In this research we directly measured the user's cognitive load to assess its effect on performance. Cognitive load can be assessed with various measures and instruments, pupillometry is one of them (Paas et al., 2010; Sweller et al., 2011). This physiological construct can be measured using eye tracking technology, which has been demonstrated to be an accurate way to measure the pupil dilation (Wang, 2011). Cognitive load can also be accessed through self-reported measurement scales such as the NASA Task Load Index (NASA-TLX). Developed for the aviation industry, the NASA-TLX is the most commonly used questionnaire-based method to assess a user's cognitive load (Cain, 2007).

Hypothesis Development

The preceding literature review shows a gap in research on the effect of information representation on certified flight instructor performance. To explore this relationship, we argue that visual attention analysis of the eye fixation count and the duration of each information representation should provide indications to the type of information representation containing the most relevant information (Holmqvist et al., 2011) and thus, indicate which type of representation fits the evaluation task best.

In a context where information representations are simultaneously presented to users, we argue that visual patterns associated with better performance should provide evidence of a cognitive fit. To analyses the visual patterns of the users, we used the scanpath which is defined as: " the route of oculomotor events through space (e.g. the first fixation in each area of interest) within a certain timespan (e.g. a trial)"(Holmqvist *et al.*, 2011, p. 254).

Past research on cognitive fit theory has mainly measured task efficiency (response time) and task effectiveness (response accuracy) during the evaluation of performances (Bacic, 2014; Vessey, 2006). Both of these measures were used in our analyses. Alternative variables such as cognitive load have been measured only in a small number of studies on the cognitive fit theory (Bacic, 2014; Shen et al., 2012). As mentioned by Bacic (2014, p. 40), prior research has "failed to actually (i) measure the impact of data representation on cognitive effort and (ii) assess the impact of users' cognitive effort on decision making efficiency and effectiveness". In this study, cognitive load serves as an alternative variable to evaluate the CFIs performance. High cognitive load decreases task performance (Sweller et al., 2011); hence information representation should facilitate problem solving effort of a user and not increase the complexity of the evaluation task.

In this research, we developed an experimental paradigm which aims to analyze the effect of simultaneously presented dynamic information representations on a single interface. In the context of our study, the evaluation task includes judging a student pilot's performance. Following Vessey and Galletta principles (1991), we argue that since this evaluation task is a spatial task, it is best supported by spatial representations. Thus, visual attention to spatial representations (3D representation, 2D representation and flight control) should lead to a better performance than attention to symbolic representations (Flight instrument and information table). *Figure 2* shows the five types of representations.

Prior research has supported the cognitive theory by demonstrating that when spatial tasks are illustrated by spatial representations, task performance is more effective (e.g. Chan et al., 2012). We thus postulate the following hypothesis.

H1. The type of information representation used will affect the accuracy of the CFI. CFIs judge more accurately the student pilot's performance when fixations on the spatial representation occur more often and for longer periods of time than the symbolic representations.

Prior research has supported the cognitive theory in demonstrating that when the task type matches the presentation format, more efficient task performance occurs (e.g. Teets et al., 2010). We thus postulate the following hypothesis.

H2. The type of information representation used should affect the evaluation time of the CFI. A shorter evaluation time should be required when the fixations on spatial representations occurs more often and for longer periods of time than when the fixations are directed at the symbolic representations.

Bacic (2014) demonstrated, using fixation duration and fixation count, that a user's cognitive effort increase when the task type and representation does not match. In an evaluation task, the nature of the information representation and the information transmission have an impact on a subject's cognitive load (Sweller et al., 2011). In aviation, prior research assessed cognitive load influences on flight performance using self-reported cognitive load and pupil dilation (e.g., Hoepf and al., 2015). In a controlled environment where ambient light is kept constant during the experiment, pupil dilation responses to cognitive load (Holmqvist et al., 2011). In the aviation field, Hoeph and his colleagues (2015) demonstrate that high cognitive demand increases the pupil diameter while low cognitive demand decreases the pupil diameter. We postulate that information representation that lead to a decreased mental cognitive load can be seen as increasing the cognitive fit. We thus formulate the following hypothesis.

H3. The type of information representation used affect the cognitive load of the CFI.

> **H3a.** A smaller pupil dilation is elicited when fixations on the spatial representation occur more often and for longer periods of time than when fixations are directed at the symbolic representation for the whole trial.

H3b. A lower self-reported cognitive load is reported when fixations on the spatial representations occur more often and for longer periods of time than fixations on the symbolic representation.

Finally, prior research in aviation has demonstrated that models of visual attention can be used to predict the human-computer interaction performance (Steelman et al.,2011; Wickens et al., 2003). Based on Venkatraman (1989, p. 432) we suggest that "frequently recurring clusters of attributes or gestalt could be found". A recurring cluster of attributes could be associated with better performance. The perspective of gestalts is defined "in terms of the degree of internal coherence among a set of theoretical attributes" (Venkatraman, 1989, p. 432), Building upon this conceptual framework, we hypothesize that perspective of fit as gestalt is the most relevant for our context. As mentioned by Venkatraman (1989), the nature of the internal congruence (visual scan patterns) among a set of strategic variables (e.g. visual metrics as the time to the first fixation in an area of interest) differs across low and high performance. Prior research has used clustering techniques to detect certain patterns of eye fixations during information processing (Mital et al., 2011). We postulate that different visual scan patterns affect the performances. We thus formulate the following hypotheses.

H4. CFIs different visual scan patterns of information representations affect their performances.

H4a. CFIs different visual scan patterns of information representations affect their accuracy.

H4b. CFIs different visual scan patterns of information representations affect their evaluation time.

H4c. CFIs different visual scan patterns of information representations affect their pupil dilation.

H4d. CFIs different visual scan patterns of information representations affect self-reported cognitive load.

Research Method

To test our hypotheses, we conducted an experiment in which a one-factor within-subjects experimental design was used to evaluate a series of landing and takeoff manoeuvres. The execution quality of the manoeuvres was manipulated. Ten participants were selected for this laboratory study. The experiment was approved by the ethics committee of our institution.

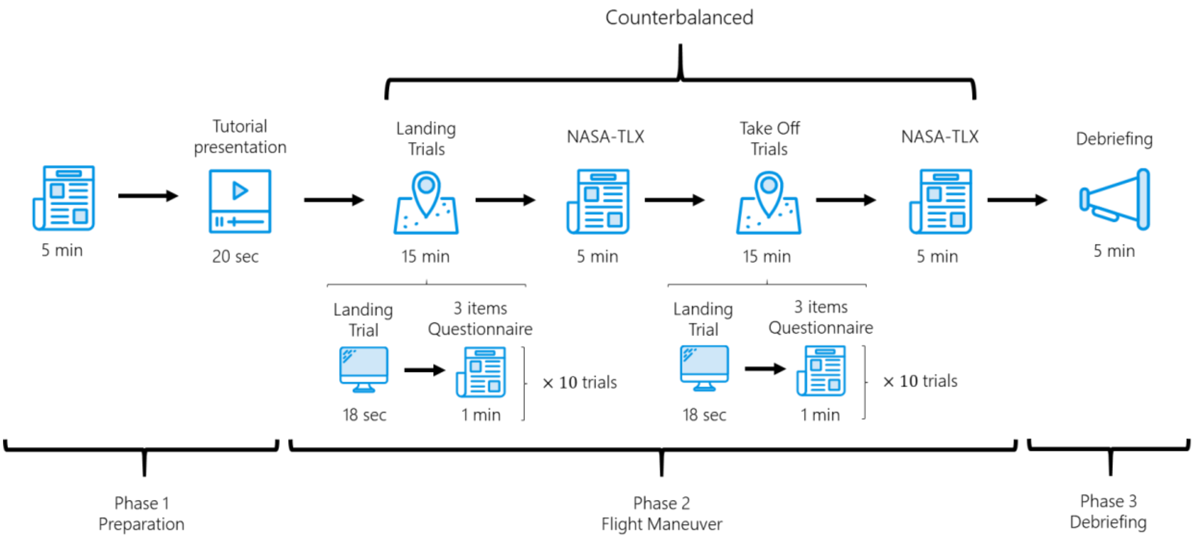
Sample

Potential participants were preselected by an aerospace company. Selection criteria was based on their expertise, and recruitment was done by the university's research laboratory via email. Compensation for their participation was their usual hourly pay rate from their aerospace company. The private aerospace company was not made aware of who participated or not in the experiment. Each participant read and signed a consent form. Participants had to be over 18 years of age, and had to have extensive experience with the interface used in the experiment. Our ten experts were either flight instructors or test engineers. The experts had minimal flight experience (mean 3171 hours of flight, median 575 hours of flight) and basic experience as a flight instructor (mean 720 hours, median 325 hours). Participants were excluded from the study if they had any of the following conditions: skin sensitivities, a cardiac pacemaker, laser-corrected vision, an inability to work at a computer without glasses, astigmatism, epilepsy, a neurological diagnosis, or any other health related diagnosis impairing an individual's ability to provide valid and normal biometric data.

Experimental Design and Procedure

Figure 1 represents the experiment's timeline. The experiment consisted of three main phases; (1) the preparation phase, (2) the flight manoeuvre phase, and (3) the debriefing phase.

Figure 1. Experimental Design



Phase 1 – Preparation

The preparation phase included the calibration of the eye-tracker and the gathering of demographic information. In preparation for the main task, participants had to watch and assess an 18-second landing manoeuvre video as a practice trial. The presentation of the video aimed to familiarize participants with the configuration of available information on the screen and with the questions asked after each trial.

Phase 2 – Flight Manoeuvre

Participants had to evaluate short flight video manoeuvres. The first task included two blocks of ten videos each. The first block figured ten videos of a flight student pilot executing a landing manoeuvre (18 seconds) or ten videos of a flight student pilot executing a takeoff manoeuvre (18 seconds). The order of the two blocks was counterbalanced. The videos include five types of representations associated either with symbolic representations (Flight instrument and information table) or spatial representations (3D representation, 2D representation and flight control). *Figure 2* shows the five types of representations.

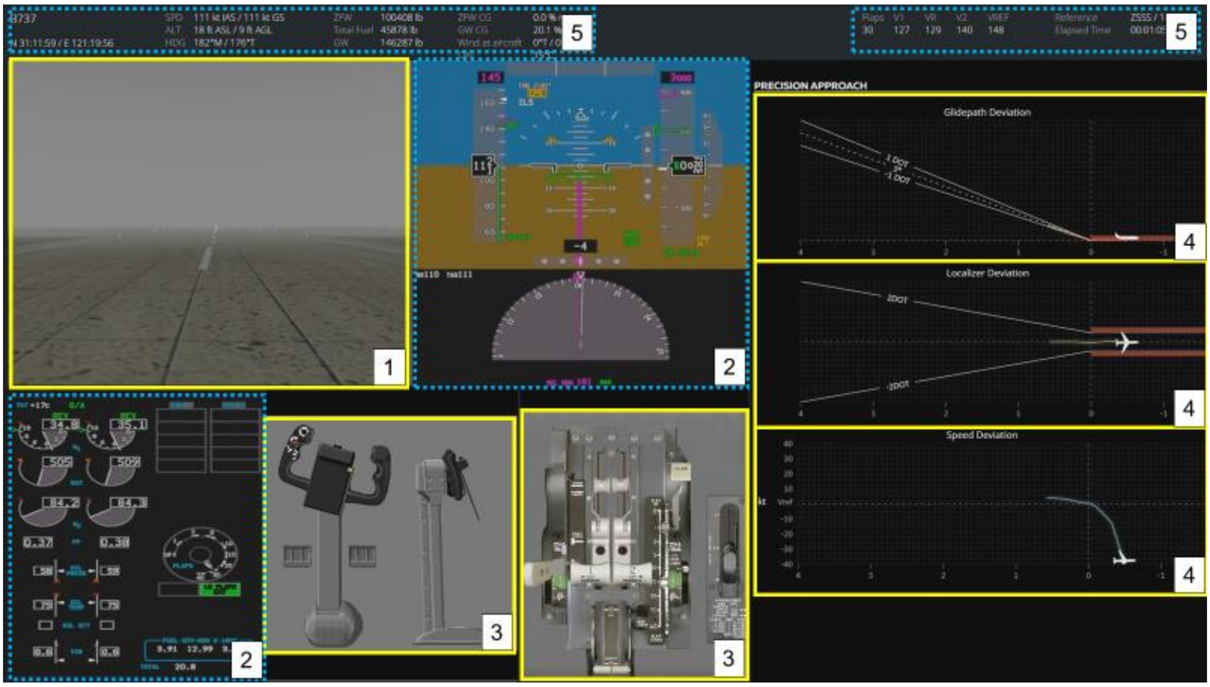


Figure 2. Presentation of stimuli - *The Display Screen Used in the Experiment Was Split into Different Areas of Interest Based on the Information Representation's Type: (1) 3D Representation; (2) Flight Instrument; (3) Flight Control; (4) 2D Representation; (5) Information Table. Symbolic Representation are Surrounded by a Blue Dotted Line, Spatial Representations are Surrounded by a Yellow Solid Line. Reproduced with the permission of the commercial research partner.*

The videos were recorded in a full-flight simulator with an expert pilot. The manoeuvres were performed on a Boeing 737. For each manoeuvre, video footage of approximately two minutes were extracted from the full-flight simulation. For each of the extracted videos, an expert pilot selected a continuous 18 second sequence judged to be the fairest representation of the student pilot's performance. The selection criteria for the landing and takeoff scenes were that they are part of the normal training process for pilots, that they are ecologically valid, and that they are short enough to accommodate the research design. The objective of having ecologically valid video sequences is to allow the CFIs to have enough information about the manoeuvre to evaluate the performance of the student pilots.

The execution quality of the manoeuvres (landing and takeoff) range from low (e.g., does not respect the standards) to high (e.g., complies with all standards). All of the landing and takeoff video sequences included multiple risk factors, such as the time of day (day or night), the presence of crosswinds, hard landings, precipitation, degree visibility (e. g., fog), centreline control, aircraft rotation, etc. Half of the trials had to be successful (a score of 3 or 4 on a 4point scale), and half had to be compromised enough to receive a low grade (a score of 1 or 2). At the end of each video, participants had to fill in a short 3-items questionnaire to evaluate the manoeuvre of the pilot student. The questionnaire was the same after each trial and consisted of three questions, which were randomized at each trial to limit any biases related to order effect.

Phase 3 - Debriefing

After the experiment, the researcher conducted a short interview to assess the participant's global impression of the experiment, their use of the different types of visualization, and their impression about the elements available on the screen in order to make a better decision about the student pilot's behaviour.

Apparatus

To measure subject's visual attention, we used a Tobii-X60 (Stockholm, Sweden), recording at a sampling frequency of 60 Hz. The eye tracker was installed under the computer screen approximately 65 centimetres away from the participant to measure eye positions, eye movements and pupil dilation after thorough calibration. Noldus Observer XT (Wageningen, Netherlands) was used to synchronize all apparatus and event markers.

Measures

Of the 10 participants that were recruited, technical issues almost completely compromised the validity of the data collection of the first 4 participants, compromised half of the data collection of 2 more, which left 4 participants with completely valid data for the analysis.

Explicit Performance Measures

Explicit student pilot's performance evaluation was measured through a 3-items questionnaire following each video trial, with these three questions being presented in random order each trial. *Table 3* shows the exact wording of these questions.

Table 3. Questions following each video trial

Question	Scale	Source
How mentally demanding was the task?	1 to 21	NASA TLX (Hart & Staveland, 1988)
How do you evaluate the student's performance in this manoeuvre?	1 to 4	Standard performance evaluation (Transport Canada, 2016)
How risky was the performance of the student?	1 to 100	Developed for this study

The 3-items questionnaire, used after each trial, includes three questions: (1) a self-reported cognitive load measured through an abridged version (1 item) of the NASA-TLX (Hart & Staveland, 1988); (2) an explicit evaluation of the student pilot's risk-taking; (3) an explicit evaluation of the student performance.

Explicit evaluation of the student pilot performance was measured by comparing the accuracy of the evaluation of each participant with the evaluation of an expert panel. The expert panel consisted of the research team member and three subject matter experts (SME). They conducted evaluations manoeuvres' quality, to create a standard evaluation score of each manoeuvre. Each SME had to evaluate the student's performance in each manoeuvre on 4-point rating scale (failed, bad, good, or excellent) and later came to an agreement on trials where their judgment was not similar. Hence, judgment accuracy of each flight manoeuvre performance was measured and was calculated by determining if the trial evaluation match the expert evaluation.

In addition, complete versions of the NASA-TLX (Hart & Staveland, 1988) and the SART (R. M. Taylor, 2017) were used after each block of trials to assess participants' mental workload.

Implicit Measures

To quantify the influence of information representation, the number and duration of the fixation on specific areas were recorded. Our study considers these metrics as an indicator of the saccadic goal, since saccades are directed to target containing relevant information (Kowler, 2011).

The display screen was split into different areas of interest (AOI) based on the information representation's type, namely spatial or symbolic representation. As illustrated in *Table 4*, the display screen was split into five AOIs. Three areas of interest correspond to spatial representations, representing 68,44% of the surface. Two areas of interest correspond to symbolic representations, representing 31,56% of the surface.

Table 4.Classification of the Area of Interest

Presentation format	Area of Interest	
	Number	Description
Spatial	1	Flight Control
Symbolic	2	Flight Instrument
Spatial	3	View 2D
Spatial	4	View 3D
Symbolic	5	Information Table

To measure the relative use of a type of representation over another, a fixation ratio was computed. The fixation ratio computes the relative use of symbolic representations over spatial representations. To calculate the fixation ratios, fixation count and duration of symbolic representation were divided by the fixation count and duration of the spatial representation, respectively, while weighing each type of representation by the number of pixels on the screen they represented. As a result of this computation, the higher the ratio, the more the participant's visual attention is on symbolic representation during a given trial.

The operationalization of the implicit variable is presented in *Table 5*. To analyze multiple psychophysiological data for a subject in each condition, the means per area of interest per trial per participant was used.

Table 5. Operationalization of the implicit variables

Construct	Instruments	Variables	Features
Cognitive load	Tobii	Pupil dilatation	Means: Means of value per AOI, for a subject in each trial.
Visual Attention	Tobii	Fixation duration and fixation count	Mean: Sum of value for a subject in each trial Weighting: weighing each AOI according to their number of pixels Ratio AOI: symbolic representation on spatial representations

		Time to first fixation	The time to the first fixation inside an area of interest during a trial
Evaluation Time	Qualtrics Survey	Time	Time: seconds

The participant decision time for each performance evaluation was measured using Qualtrics Reaction Time Engine (QRTEngine) included in Qualtrics Survey. The decision time refers to the time each participant took to answer the 3-item questionnaire following each video trial.

Results

Fixation Ratios

To test H1, H2 and H3, multiple models of regression analysis were used. Logistic regression was used to test if the fixation ratio significantly predicted participants' accuracy (H1). Linear regression with mixed model (20 trials for each participants) and two tailed levels of significance was used to test if the effect of the fixation ratios over the participants' evaluation time (H2), the participants pupil dilation (H3a) and the participant's self-perceived cognitive load (H3b).

Table 6. Fixation Ratio Results

	Accuracy (H1)	Evaluation Time (H2)	Self-Reported Cognitive Load (H3a)	Pupil Dilation (H3b)
	Coef. (Std.Err.)	Coef. (Std.Err.)	Coef. (Std.Err.)	Coef. (Std.Err.)
Ratio Fixation Count	-0.504 (0.433)	-0.887 (1.181)	0.227 (0.370)	-0.011 (0.018)
Ratio Fixation Duration	0.331 (0.415)	0.661 (1.161)	-0.455 (0.361)	0.027 (0.018)
Note: ** p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001				

Results indicated that the effect of the fixation count ratios over participants' accuracy (-0.0504, p=0.248) (H1), evaluation time (-0.887, p=0.455) (H2) and self-reported cognitive load (0.227, p=0.541) (H3a) and pupil dilation (-0.011, p=0.557) (H3b) was non-significant (*Table 6*). Results indicated that the effect of the fixation duration ratios over participants' accuracy

(0.331, p=0.428) (H1), evaluation time (0.661, p=0.571) (H2) and self-reported cognitive load (-0.455, p=0.211) (H3a) and pupil dilation (0.027, p=0.134) (H3b) was non-significant (*Table 6*). H1, H2 and H3 were not supported.

Visual Attention Pattern

To test H4, a curve clustering technique was applied (Abraham, 2003). While taking into account the temporal and dynamic nature of the task, this technique was used to test the "fit as gestalt" proposed by Venkatraman (1989). A curve was made based on the order of the participants' visual sequence (specifically the time to the first fixation in each area of interest), for each trial of each participant. Then, a distance-based curve clustering algorithm was applied (Montero and Vilar, 2014), in which a correlation-based distance between each pair of curves was calculated. Then, the matrix of distance was put into a classical K-Means clustering algorithm to classify similar curves into the same class. As illustrated in *Table 7*, the curves were classified into 6 clusters. The sequence name refers to the order of the participants' visual sequence (See *Table 4*). The first number corresponds to the first AOI fixated while the second number to the second AOI fixated, the third number to the third AOI fixated and the fourth number to the fourth AOI fixated. A cluster name containing three number mean that only three AOI were fixated during the participants' visual sequence. A cluster name containing four number mean that four AOI were fixated during the participants' visual sequence. The cluster other includes all the other sequences that do not belong to any other clusters.

Table 7. Clusters Classification – Pairwise Comparison of the Area of Interest

Sequence Name*	Number of Observations
2-4-1	27
2-4-3	3
2-4-3-1	8
2-4-1-3	23
2-4-5-1	7
Others	11

* Each number in the sequence name correspond to a specific AOI and a specific order in which the AOI was fixated.

Linear regression with mixed model and two tailed levels of significance was used to test if CFIs different visual scan patterns of information representations affect their performances (H4). Results indicated that the pairwise comparison of the accuracy (H4a) and the self-reported cognitive load (H4d) between the different sequence clusters were non-significant.

Results indicated that the pairwise comparation of the evaluation time between the different sequence clusters (*Table 8*) was significant (H4b). Results indicated that this test was non-significant when conducted while controlling for accuracy.

Table 8. Pairwise Comparison of the Evaluation Time Between Clusters (H4c)

Sequence 1 - Name		Sequence 2 - Name	Sequence 1 - Mean	Sequence 2 - Mean	Estimate	Standard Error
241	vs	243	11.89	9.14	-13.04**	3.20
241	vs	Others	11.89	12.45	-1.95	2.19
241	vs	2413	11.89	15.13	-3.24	1.62
241	vs	2431	11.89	10.25	1.90	2.09
241	vs	2451	11.89	-0.29	-6.18	2.22
243	vs	Others	9.14	12.45	11.10*	3.38
243	vs	2413	9.14	15.13	9.80	3.09
243	vs	2431	9.14	10.25	14.94**	3.41
243	vs	2451	9.14	-0.29	6.86	3.46
2413	vs	2431	15.13	10.25	5.14	2.13
2413	vs	2451	15.13	-0.29	-2.94	2.24
2431	vs	2451	10.25	-0.29	-8.08	2.57
Others	vs	2413	12.45	15.13	-1.30	2.04
Others	vs	2431	12.45	10.25	3.84	2.56
Others	vs	2451	12.45	-0.29	-4.24	2.64

*Note: ** p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001*

Results indicated that the average decision time in the sequence 241 (flight instrument, 3D view and flight control) was significantly lower than in the sequence 243 (flight instrument, 3D view and 2D view). The effect size (d = -0.99) for this analysis showed a large effect according to Cohens (1998) convention (d = .80). Furthermore, the results showed that the average decision time in the sequence 243 (flight instruments, 3D view and 2D view) was significantly higher than in the sequence 2431 (flight instrument, 3D view, 2D view and flight control). The effect size (d = 1.06) for this analysis showed a large effect according to Cohens (1998) convention (d = .80). H4b is supported, information representation lead to different visual patterns of the CFIs which will affect their evaluation time. Results indicated that this test was non-significant when conducted while controlling for accuracy.

Results indicated that the pairwise comparation of the pupil dilation between the different sequence clusters (*Table 9*) was significant (H4c).

Table 9. Pairwise Comparison of the Pupil Dilation Between Clusters (H4d)

Sequence 1 - Name		Sequence 2 - Name	Sequence 1 - Mean	Sequence 2 - Mean	Estimate	Standard Error
241	vs	243	3.73	3.79	-0.18*	0.06
241	vs	Others	3.73	3.65	-0.08	0.03
241	vs	2413	3.73	3.62	-0.07	0.03
241	vs	2431	3.73	3.73	0.00	0.04
241	vs	2451	3.73	3.79	-0.05	0.04
243	vs	Others	3.79	3.65	0.11	0.06
243	vs	2413	3.79	3.62	0.17	0.06
243	vs	2431	3.79	3.73	0.19*	0.06
243	vs	2451	3.79	3.79	0.13	0.06
2413	vs	2431	3.62	3.73	0.07	0.04
2413	vs	2451	3.62	3.79	0.02	0.04
2431	vs	2451	3.73	3.79	-0.05	0.04
Others	vs	2413	3.65	3.62	0.01	0.03
Others	vs	2431	3.65	3.73	0.08	0.04
Others	vs	2451	3.65	3.79	0.03	0.04

*Note: ** p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001*

Results indicated that the average pupil dilation in the sequence 241 (flight instrument, 3D view, flight control) was significantly lower than in the sequence 243 (flight instrument, 3D view, 2D view). The effect size (d = -0,36) for this analysis showed a small effect according to Cohens (1998) convention (d = .20). The results reported that the average pupil dilation in the sequence 243 (flight instruments, 3D view, 2D view) was significantly higher than in the sequence 2431 (flight instrument, 3D view, 2D view, flight control). The effect size (d = 0,36) for this analysis showed a small effect according to Cohens (1998) convention (d = .20). H4c is supported, information representation lead to different visual patterns of the CFIs which will affect their pupil dilation. Results indicated that this test was non-significant when conducted while controlling for accuracy.

Discussion

Summary of the results

The objective of this exploratory study was to investigate the extent to which the cognitive fit between simultaneously presented information representations on a CFIs dashboard and an evaluation task influence the CFIs performance in a simulator training context. First, results showed a non-significant relationship between fixations ratios and the dependant variables (H1, H2, H3). Results showed a non-significant relationship between fixations ratios (H1) and accuracy; between fixations ratios and the evaluation time (H2); between fixations ratios and the pupil dilation (H3a) and between fixations ratios and self-perceived cognitive load (H3). Considering the nature of the task, one would expect the results to be significant. A large amount of studies supporting the cognitive fit theory have use time and accuracy to measure the performance in complex spatial task (Bacic, 2014; Speier et al. 2003; Teets et al. 2010). The nature of the task, the methodology or the statistical power of this study could explain why theses hypotheses were not supported. Second, results suggested that an ideal visual pattern lead to a more efficient task performance (H4b). Participants who employed a visual sequence that begun at the flight instrument, followed by the 3D view and then the 2D view (2-4-3), had significantly lower evaluation times than participants who first fixated the flight instrument, followed by the 3D view and finally the flight control. Result suggest that fixating the 2D view first increases evaluation time. Participants' whose visual sequence began at the flight instrument, followed by the 3D view, the 2D view and ended with the flight control (2-4-3-1), had significantly higher evaluation times than when they first fixated the flight instrument, followed by the 3D view and the 2D view. Results suggest that adding the flight control to the visual sequence decreases the evaluation time of the participants. Third, the results suggest that visual pattern affects task performance in terms of cognitive load (H4c). The results indicate that average pupil dilation for visual sequences that begin at the flight instrument, followed by the 3D view, flight control (2-4-1) were significantly lower than sequences where fixations debuted at the flight instrument, followed by the 3D view and the 2D view (243). Result show that looking at the 2D view instead of the flight controls following the initial sequence increases pupil dilation. We also report that the average pupil dilation in the participant visual sequences who first looked at the flight instruments, followed by the 3D view, 2D view (2-4-3) were significantly higher than when they first fixated on the flight instrument, followed by the 3D view, the 2D view and the flight control (2-4-3-1). Results suggest that the looking at the flight control following the 2D view decrease the pupil dilation of the participants.

Theoretical Contributions

Our research contributes four main findings. First, this study is, to our knowledge, the first to investigate the influence of simultaneously presenting multiple information representations while using the framework of the cognitive fit theory (Bacic, 2014; Vessey, 2006). The unique contribution of the article lies in the dynamic method used to measure the cognitive fit. The application of a curve clustering technique, based on the order of the participants' visual sequence, was a creative and innovative methodology to evaluate the cognitive fit theory. Our application of the cognitive fit theory to the aviation field was also innovative (Bacic, 2014). We put forward a proof of concept that should allow future work to test our paradigm with a larger sample and promote other researcher to use curve clustering technique to test the cognitive fit theory.

Second, it extends previous work that studied the impact of interface displays in aviation on pilot's performance (Alexander et al., 2005; Andre et al., 1991; Haskell & Wickens, 1993; Smallman, St John, Cowen, Oonk, & Cowen, 2001; Olmos et al., 2000; Wickens, 2002) to CFIs performance. Since different representations are simultaneously presented on a flight interface, we have demonstrated the validity of using our experimental paradigm to investigate the relative costs and benefits of different aviation displays in a higher ecological validity context.

Third, we also support the cognitive fit theory (Vessey & Galletta, 1991), according to which, to determine whether one representation is better than another, an evaluation of the interaction between the information representation and the task is required; both must fit together and be directed towards the same goal. A literature review, provided by Vessey (2006), has shown that most studies on the cognitive fit theory have only accounted for time and accuracy as dependent variables when attempting to evaluate the cognitive fit. As suggested by (Bacic, 2014) our research extends this theory by adding cognitive load as a dependent variable to evaluate the fit between information representation and a complex task.

Fourth, in response to the lack of research done on CFIs (Lazure et al., 2018), the present study highlights the need to conduct further research on CFIs performance using physiological measurement tools. We argue that researchers should consider using physiological measurement tools in future research to investigate their influence over CFIs performance. The development of a better suited interface to the needs of CFIs is a first step to improving their teaching and to gaining better understanding of CFIs' performance factors.

Practical Implications

Our findings can be useful to the Human Factor Specialist in the aviation industry by developing an experimental paradigm to investigate the dynamic cognitive fit between a task and information representations (Vessey, 2006; Bacic, 2014). This paradigm could be used to test the design of aviation displays. Results also demonstrated the relevance of using biometric tools as a useful method for evaluating the design of their interface. Eye-tracking analysis allowed actionable conclusions about the interface used in the experiment. Using eye-tracking analyse, this paradigm could potentially be used to create more adaptative interactions with the interface. When the user deviates from the optimal scan pattern, participants could receive feedback from the interface, which could help the user's decision-making process. It will be possible to apply this method in future projects to answer user experience and user interface design questions, that will help the human factor specialist build more efficient interfaces for their instructors and pilots.

Second, since instructors represent a key element in the development of aircraft pilots, increasing the quality of the CFIs teaching by enabling them to make more efficient and effective judgments might improve the effectiveness of the simulation training session, which could, in turn lead to the development of better pilots. CFIs gain further insight on the importance of focusing their attention on different representations, so they can make more efficient evaluations. CFIs could use our paradigm in a training session to gain further insight on their visual attention and use this information to improve their interaction with their interface.

Limitations and Further Research

This research includes several limitations that must be acknowledged. First, the number of participants is low, since they are from a very specific, small, and highly experienced population. This has a direct impact on statistical power and the ability to generalize results. Based on the previous literature on cognitive fit theory, we were expecting our hypotheses on accuracy and time to be supported (Bacic, 2014; Vessey, 2006; Teets et al., 2010). This could be explained by the simple size used in this experiment, which could not be sufficient enough. Further studies should replicate the experimentation to obtain a larger sample size and gain more statistical power. Second, the stimulus used in the practice sequence was reused in the landing block. Two out of the four participants rated the second presentation of the same stimuli as riskier than the first time they evaluated it during the practice trial. Due to the small number of trials, it is difficult to establish a statistically significant relationship between seeing the stimuli for a second time and lower evaluation of risk. Third, in this research, we consider the evaluation time as a conceptual process that includes not only the evaluation of the student pilot's performance, but also the evaluation of the student pilot's risk-taking behaviors and the evaluation of the participant's self-perceived cognitive load. Evaluation time includes the time used to answer all three questions in the questionnaires, although a single question addressed the pilot student's performance. This represents a confounding factor that limits the possibility of identifying the true time taken to evaluate the performance. The validity of our methodology could have impacted the lack of support for our hypotheses. Further studies should measure the time for the task of judging the student pilot's performance in more precise manner. In support of Teets et al. (2010), we argue that further studies should also add time pressure to the participant in order to limit their evaluation time. Visual patterns were found to influence both evaluation time and pupil dilation, yet are uncorrelated to accuracy. Therefore, participants with a strong visualization pattern, short evaluation time and minimal pupil dilation may not produce accurate evaluations of the student's performance. Future work should aim to replicate this study in order to more strongly establish the correlational relationship between these factors.

The experimental paradigm, in which multiple short video stimuli were shown to participants, induced a high variability of psycho-physiological states. In a future experiment with a larger sample size, the identification of statistically powerful significant relationships between these states and performance measures could be established.

Conclusion

In conclusion, this experiment demonstrates the feasibility of our experimental paradigm as a tool to assess the cognitive fit of simultaneously available information representations on CFIs performance. When different types of information representations are presented simultaneity on an interface, our paradigm demonstrates that visual attention can be used to investigate the combined and dynamic effect of different information representations over performance. Our results have successfully demonstrated that a better dynamic cognitive fit can be achieved via different visual scan patterns. Developing interfaces better suited for their tasks will improve the CFIs performance. Improving CFIs performance could foster the development of better and faster trained flight pilots.

Reference

Abraham, C., Cornillon, P. A., Matzner-Løber, E. R. I. C., & Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian journal of statistics*, *30*(3), 581-595.

Alexander, A. L., Wickens, C. D., & Merwin, D. H. (2005). Perspective and coplanar cockpit displays of traffic information: Implications for maneuver choice, flight safety, and mental workload. International Journal of Aviation Psychology, 15(1), 1–21.

Andre, A. D., Wickens, C. D., & Moorman, L. (1991). Display formatting techniques for improving situation awareness in the aircraft cockpit. The International Journal of Aviation Psychology, 1(3), 205-218.

Bacic, D. (2014). The Role of Cognitive Effort in Decision Performance Using Data Representations:; a Cognitive Fit Perspective.

Cain, B. (2007). A Review of the Mental Workload Literature. Defence Research and Development Toronto (Canada).

Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research, 51(13), 1484–1525.

Chan, H. C., Goswami, S., & Kim, H.-W. (2012). An Alternative Fit through Problem Representation in Cognitive Fit Theory. Journal of Database Management, 23(2), 22–43.

De Guinea, A. O., Titah, R., & Léger, P.-M. (2014). Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. Journal of Management Information Systems, 30(4), 179–210.

Dehais, Frédéric, Causse, M., & Pastor, J. (2008). Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring Frédéric. In Proceedings from the 3rd International Conference on Research in Air Transportation.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in Psychology, 52, 139–183.

Haskell, I. D., & Wickens, C. D. (1993). Two- and Three- Dimensional Displays for Aviation: A Theoretical and Empirical Comparison. The International Journal of Aviation Psychology, 3(2), 37–41.

Hoepf, M., Middendorf, M., Epling, S., & Galster, S. (2015). Physiological Indicators of Workload in a Remotely Piloted Aircraft Simulation. Air Force Research Lab Wright-Patterson AFB, OH: United States Air Force Air Material Command.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

John, T., & Kundisch, D. (2015). Creativity Through Cognitive Fit : Theory and Preliminary Evidence in a Business Model Idea Generation Context. Wirtschaftsinformatik 2015, (JANUARY 2015), 1283–1297.

Khatri, V., Vessey, I., Ram, S., & Ramesh, V. (2006). Cognitive fit between conceptual schemas and internal problem representations: The case of geospatio-temporal conceptual schema comprehension. IEEE Transactions on Professional Communication, 49(2), 109–127.

Kowler, E. (2011). Eye movements: The past 25years. Vision Research, 51(13), 1457–1483.

Lazure, C., Dumont, L., El Mouderrib, S., Delisle, J-F., Senecal, S., Leger, P.M. (2018). Evaluation of Certified flight instructors' performance factor in a flight simulation training context. Manuscript submitted for publication.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*(1), 5-24.

Montero, P., & Vilar, J. A. (2014). Tsclust: An r package for time series clustering. *Journal of Statistical Software*, *62*(1), 1-43.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. W. M. (2010). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist, 1520(38), 43–52.

Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-Tracking Measures in Aviation: A Selective Literature Review. The International Journal of Aerospace Psychology, 1-15.

Riedl, R., & Léger, P.-M. (2016). Tools in NeuroIS Research: An Overview. In S.-V. B. H. 2016 (Ed.), Fundamentals of NeuroIS (p. 115).

Shen, M., Carswell, M., Santhanam, R., & Bailey, K. (2012). Emergency management information systems: Could decision makers be supported in choosing display formats? Decision Support Systems, 52(2), 318–330.

Smallman, H. S., St John, M., Cowen, M. B., Oonk, H. M., & Cowen, M. B. (2001). Availability in 2D and 3D Displays. IEEE Computer Graphics and Applications, 21(5), 51–57.

Steelman, K. S., McCarley, J. S., & Wickens, C. D. (2011). Modeling the Control of Attention in Visual Workspaces. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(2), 142–153.

Speier, C., Vessey, I., & Valacich, J. S. 2003. The Effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision-Making Performance. Decision Sciences, 34(4): 771-797.Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory.

Olmos, O., Wickens, C. D., & Chudy, A. (2000). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering. The International Journal of Aviation Psychology, 10(3), 247-271.

Taylor, R. M. (2017). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. Situational Awareness, 111–128.

Teets, J. M., Tegarden, D. P., & Russell, R. S. (2010). Using cognitive fit theory to evaluate the effectiveness of information visualizations: An example using quality assurance data. IEEE Transactions on Visualization and Computer Graphics, 16(5), 841–853.

Transport Canada. (2016). Flight Test Guide.

Venkatraman, A. N. (1989). The Concept of Fit in Strategy Research: Toward Verbal and Statistical Correspondence. Academy of Management Review, 14(3), 423–444.

Vessey, I. (1991). Cogntive Fit: A Theory-based analysis of the Graphs versus Tables Literature. Decision Science, 22(2), 219–240.

Vessey, I. (2006). The theory of cognitive fit. Human-Computer Interaction and Management Information Systems: Foundations, 141–183.

Vessey, I., & Galletta, D. (1991). Cognitive Fit: An Empirical Study of Information Acquisition. Information System Research, 2(1), 63–84.

Vidulich, M. A., Wickens, C. D., Tsang, P. S., & Flach, J. M. (2010). Information Processing in Aviation. Human Factors in Aviation, 175–215.

Wang, J. T. Y. (2011). Pupil dilation and eye tracking. A handbook of process tracing methods for decision research: A critical review and user's guide, 185-204.

Wickens, C. D. (2002). Situation awareness and workload in aviation situation awareness and workload in aviation. Psychological Science, 11(4), 128–133.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. Human Factors: The Journal of the Human Factors and Ergonomics Society, 45(3), 360–380.

Yu, C. S., Wang, E. M. Y., Li, W. C., & Braithwaite, G. (2014). Pilots' visual scan patterns and situation awareness in flight operations. Aviation Space and Environmental Medicine, 85(7), 708–714.

Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. International Journal of Aviation Psychology, 26(3–4), 75–104. 6

Chapitre 4: Conclusion

Ce mémoire avait pour objectifs de faire l'état des connaissances académiques sur les instructeurs de vol et de comprendre, en se basant sur la théorie de l'adéquation cognitive, l'effet des représentations dynamiques de l'information sur la performance des instructeurs de vol. Dans un premier temps, ce mémoire a réalisé une revue de littérature systématique sur les instructeurs de vol et réalisé un examen de portée (*scoping review*) sur les facteurs de performance des instructeurs de vol en contexte de formation en simulateur. En second lieu, ce mémoire a étudié l'influence des représentations dynamiques de l'information sur la prise de décision des instructeurs de vol en contexte de formation en simulateur. Grâce à l'utilisation des outils biométriques, ce mémoire a développé un paradigme expérimental qui permet d'étudier dans quelles mesures une adéquation cognitive, entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur. Ce paradigme est innovant, car il permet d'étudier l'effet des représentations dynamiques de l'information sur les performances des utilisateurs en utilisant différents modèles de visualisation. Ainsi, ce paradigme permet de comparer différentes représentations dynamiques de l'information entre elles pour savoir laquelle ou lesquelles sont le plus adaptées à une tâche particulière.

Une étude en laboratoire a permis de récolter des mesures auto-rapportées (explicites) ainsi que des mesures comportementales et biométriques (implicites). Cette expérience a été réalisée auprès de 10 participants experts dont 4 participants avec des données totalement valides pour l'analyse. Les participants devaient évaluer un pilote de vol réalisé une série de manœuvres d'atterrissage et de décollage.

Nous profiterons de ce chapitre pour rappeler les questions de recherche, les hypothèses et la méthodologie nous ayant permis de conduire cette expérience. Nous récapitulerons les résultats et la contribution des deux articles. Enfin, nous exposerons les limites globales du mémoire et les recommandations pour de futures recherches.

Rappel des questions de recherche et de la méthodologie

Afin d'obtenir une meilleure compréhension des facteurs pouvant influer sur les performances des instructeurs de vol, un état des connaissances académique a été réalisé. Cette recherche répond à la question suivante :

Question 1 : *Quels facteurs influencent l'état psychophysiologique des instructeurs de vol dans un contexte de formation en simulateur?*

Puisque les instructeurs de vol ont recours à une interface pour évaluer la performance des pilotes de vol, l'auteur de ce mémoire a également réalisé une expérience pour étudier l'influence des représentations dynamiques de l'information sur la performance. Ce mémoire fournit un article de type exploratoire pour répondre à la question de recherche ci-dessous :

Question 2 : *Dans un contexte où les représentations de l'information sont présentées simultanément à un utilisateur, est-ce qu'il y a des modèles de visualisation associés à de meilleures performances?*

Afin de répondre à cette question de recherche, l'analyse de l'attention visuelle a fourni les indications nécessaires permettant de comprendre l'effet de différents modèles de visualisation sur la performance. Les mesures de fixations de chaque participant envers les représentations de l'information ont permis d'évaluer quel type de représentation était le mieux adapté à la tâche. Pour ce faire, la performance a été évaluée selon trois aspects : l'exactitude, le temps et la charge cognitive. Nos hypothèses explorent comment l'attention visuelle peut avoir une influence sur l'exactitude de leur évaluation, le temps d'évaluation et la charge cognitive des instructeurs. Ainsi, à travers cet article les 4 hypothèses suivantes sont présentées ci-dessous :

H1 - Le type de représentation d'informations utilisé affecte l'exactitude des évaluations réalisées par les instructeurs de vol. Les instructeurs de vol évaluent plus précisément la performance de l'étudiant pilote lorsque les fixations sur les représentations spatiales sont plus fréquentes et de plus longue durée que les représentations symboliques.

H2 - Le type de représentation d'informations utilisé affecte le temps d'évaluation des instructeurs de vol. Un temps d'évaluation plus court est requis lorsque les fixations sur les représentations spatiales sont plus fréquentes et de plus longue durée que les représentations symboliques.

H3 - Le type de représentation d'informations utilisé affecte la charge cognitive des instructeurs de vol.

> **H3a - Une dilatation de la pupille moins importante est mesurée lorsque les fixations sur les représentations spatiales se produisent plus souvent et pendant de plus longues périodes que les représentations symboliques.**
>
> **H3b - Une charge cognitive perçue moins importante est mesurée lorsque les fixations sur les représentations spatiales se produisent plus souvent et pendant de plus longues périodes que les représentations symboliques.**

H4 - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent leurs performances.

> **H4a - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent l'exactitude des évaluations réalisées par les instructeurs de vol.**
>
> **H4b - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent le temps d'évaluation.**
>
> **H4c – Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent la dilatation de la pupille.**
>
> **H4d - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent la charge cognitive perçue.**

Au niveau de la méthodologie, l'expérience avait une durée d'environ 90 minutes. L'objectif était de comprendre l'effet des représentations dynamiques de l'information, lorsque celles-ci sont présentées simultanément, sur les performances des instructeurs de vol. Afin de relier les mesures implicites (réactions cognitives) aux mesures explique (évaluation d'une manœuvre), nous avons utilisé différents modèles de visualisation. L'effet des modèles de visualisation a été analysé grâce au logiciel statistique SAS.

Principaux résultats

Afin de répondre à notre première question de recherche, une revue de littérature systématique a permis d'identifier seulement 20 articles s'étant intéressés directement aux instructeurs de vol depuis 1965. Par la suite, une revue de la portée (*scoping review*) a été réalisée sur les facteurs de performance des instructeurs de vol en contexte de formation en simulateur. Les résultats de cette revue ont permis d'identifier plusieurs construits à considérer lors de l'évaluation de la performance des instructeurs de vol œuvrant dans un contexte de formation en simulateur, spécifiquement : la conscience situationnelle, l'attention, la détection des changements, la tolérance aux risques, le flow et l'absorption cognitive, la charge cognitive et le stress.

Les résultats de notre expérience en laboratoire ont permis de répondre la deuxième question de recherche du mémoire. Les résultats sont présentés selon les hypothèses.

H1 - Le type de représentation d'informations utilisé affecte l'exactitude des évaluations réalisées par les instructeurs de vol. Les instructeurs de vol évaluent plus précisément la performance de l'étudiant pilote lorsque les fixations sur les représentations

spatiales sont plus fréquentes et de plus longue durée que les représentations symboliques.

Les résultats sont non significatifs. Le type de représentation d'informations utilisé n'a pas d'impact sur l'exactitude des évaluations réalisées par les instructeurs de vol.

H2 - Le type de représentation d'informations utilisé affecte le temps d'évaluation des instructeurs de vol. Un temps d'évaluation plus court est requis lorsque les fixations sur les représentations spatiales sont plus fréquentes et de plus longue durée que les représentations symboliques.

Les résultats sont non significatifs. Le type de représentation d'informations utilisé n'a pas d'impact le temps d'évaluation des instructeurs de vol.

H3 - Le type de représentation d'informations utilisé affecte la charge cognitive des instructeurs de vol.

Les résultats sont non significatifs. Le type de représentation d'informations utilisé n'a pas d'impact la charge cognitive des instructeurs de vol.

> **H3a - Une dilatation de la pupille moins importante est mesurée lorsque les fixations sur les représentations spatiales se produisent plus souvent et pendant de plus longues périodes que les représentations symboliques.**
>
> Les résultats sont non significatifs. Le type de représentation d'informations utilisé n'a pas d'impact la dilatation de la pupille des instructeurs de vol.
>
> **H3b - Une charge cognitive perçue moins importante est mesurée lorsque les fixations sur les représentations spatiales se produisent plus souvent et pendant de plus longues périodes que les représentations symboliques.**
>
> Les résultats sont non significatifs. Le type de représentation d'informations utilisé n'a pas d'impact la charge cognitive perçue des instructeurs de vol.

H4 - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent leurs performances.

L'hypothèse H4 est partiellement soutenue. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent leurs performances au niveau du temps d'évaluation et de la dilatation de la pupille.

> **H4a - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent l'exactitude des évaluations réalisées par les instructeurs de vol.**

Les résultats sont non significatifs. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations n'ont pas d'impact sur l'exactitude des évaluations.

H4b - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent le temps d'évaluation.

Les résultats sont significatifs. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations ont un impact sur le temps d'évaluation. Une analyse des modèles de visualisation des participants permet d'observer un effet sur le temps d'évaluation entre différentes séquences de visionnement.

Par rapport à la séquence de visionnement débutant par les instruments de vol, suivis de la représentation en trois dimensions et des représentations en deux dimensions, le temps d'évaluation des participants était nettement plus court que lorsqu'ils avaient d'abord fixé les instruments de vol, suivis de la représentation en trois dimensions et de la représentation des commandes de vol. Les résultats suggèrent que les représentations en deux dimensions augmentent le temps d'évaluation des participants. De plus, comparativement à la séquence de visionnement débutant par les instruments de vol, suivis de la représentation en trois dimensions, des représentations en deux dimensions et de la représentation des commandes de vol, le temps d'évaluation des participants était significativement plus long lorsqu'ils avaient d'abord fixé les instruments de vol, suivis de la représentation en trois dimensions et des représentations en deux dimensions. Les résultats suggèrent que l'ajout de la représentation des commandes de vol à la séquence visuelle réduit le temps d'évaluation des participants.

H4c – Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent la dilatation de la pupille.

Les résultats sont significatifs. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations ont un impact sur la dilatation de la pupille. Une analyse des modèles de visualisation des participants permet d'observer un effet sur la dilatation de la pupille entre différentes séquences de visionnement.

Par rapport à la séquence de visionnement débutant par les instruments de vol, suivis de la représentation en trois dimensions et la représentation des commandes de vol, la dilatation de la pupille des participants était moins élevée que lorsqu'ils avaient d'abord fixé les instruments de vol, suivis de la représentation en trois dimensions et des représentations en deux dimensions. Les résultats suggèrent que les représentations en deux dimensions engendrent une plus grande dilatation de la pupille. De plus, comparativement à la séquence de visionnement débutant par les instruments de vol, suivis de la représentation en trois dimensions et des représentations en deux dimensions, la dilatation de la pupille des participants était significativement plus élevée que lorsqu'ils avaient fixé au départ les instruments de vol, suivis de représentation en trois dimensions, des représentations en deux dimensions et de la représentation des commandes de vol. Les résultats suggèrent que l'ajout de la représentation des commandes de vol à la séquence visuelle augmente la dilatation de la pupille.

H4d - Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent la charge cognitive perçue.

Les résultats sont non significatifs. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations n’ont pas impact sur la charge cognitive perçue.

La *Table 10* présente un sommaire des hypothèses supportées et non supportées de la seconde question de recherche du mémoire.

Table 10. Tableau Sommaire des Hypothèses

Hypothèses		Résultats
H1.	Le type de représentation d'informations utilisé affecte l'exactitude des évaluations réalisées par les instructeurs de vol. Les instructeurs de vol évaluent plus précisément la performance de l'étudiant pilote lorsque les fixations sur les représentations spatiales sont plus fréquentes et de plus longue durée que les représentations symboliques.	Non supportée
H2.	Le type de représentation d'informations utilisé affecte le temps d'évaluation des instructeurs de vol. Un temps d'évaluation plus court est requis lorsque les fixations sur les représentations spatiales sont plus fréquentes et de plus longue durée que les représentations symboliques.	Non supportée
H3.	Le type de représentation d'informations utilisé affecte la charge cognitive des instructeurs de vol.	Non supportée
	H3a. Une dilatation de la pupille moins importante est mesurée lorsque les fixations sur les représentations spatiales se produisent plus souvent et pendant de plus longues périodes que les représentations symboliques.	Non supportée
	H3b. Une charge cognitive perçue moins importante est mesurée lorsque les fixations sur les représentations spatiales se produisent plus souvent et pendant de plus longues périodes que les représentations symboliques.	Non supportée
H4.	Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent leurs performances.	Partiellement supportée
	H4a. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent le temps d'évaluation.	Non supportée
	H4b. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent le temps d'évaluation.	Supportée
	H4c. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent la dilatation de la pupille.	Supportée
	H4d. Différents modèles de visualisation des instructeurs de vol envers les représentations d'informations affectent la charge cognitive perçue.	Non supportée

Contributions du mémoire

Plusieurs contributions, tant au niveau théorique que pratique, découlent de ce mémoire. Grâce aux articles, ce mémoire contribue à combler le manque dans la littérature sur les instructeurs de vol dans un contexte de formation en simulateur. Ce premier article propose de considérer une série de construits attribuable à la psychologie et la physiologie pour évaluer la performance des instructeurs. Les chercheurs pourraient tirer profit de notre recherche, car elle fournit une liste de mesures validées par la littérature qui pourraient être utilisées dans des expériences futures pour évaluer les performances des instructeurs de vol en contexte de formation en simulateur. L'utilisation conjointe d'outils psychométriques et biométriques pourrait fournir une compréhension claire du processus décisionnel et attentionnel des instructeurs de vol.

L'expérience en laboratoire contribue également aux chercheurs. Notre recherche apporte trois contributions principales. Premièrement, ce mémoire expose la validité de recourir à notre paradigme expérimental, soit l'utilisation d'une technique statistique de recouvrement des courbes, pour étudier les avantages et les inconvénients des représentations dynamiques de l'information. En le comparant à la littérature existante (e.g. Wickens et al., 2000; Alexander et al., 2005; Shen *et al.*, 2012), nous pensons que cette méthode pourrait potentiellement conduire à des résultats constants à travers des études futures.

Deuxièmement, ce mémoire supporte la théorie de l'adéquation cognitive (Vessey, 1991) selon laquelle, pour déterminer la supériorité d'une représentation par rapport à une autre, l'évaluation de l'interaction entre la représentation de l'information et la tâche est requise; les deux doivent viser le même objectif. Vessey (2006) a observé que la grande majorité des études sur la théorie de l'adéquation cognitive évalue la performance selon deux principaux aspects : le temps et l'exactitude. Tel que suggéré par (Bacic, 2014), ce mémoire étend la théorie de l'adéquation cognitive en considérant l'étude de la charge cognitive pour évaluer les avantages et les inconvénients des différents types de représentations visuels. Cette étude est la première à évaluer l'influence des représentations de l'information selon une analyse non séquentielle et étendre la théorie de l'adéquation cognitive au domaine de l'aviation (Vessey, 2006; Bacic, 2014).

Troisièmement, ce mémoire contribue aux recherches en expérience utilisateur. Le paradigme expérimental développé par ce mémoire utilise différents modèles de visualisation, pour étudier dans quelles mesures une adéquation cognitive, entre une tâche d'évaluation et des représentations dynamiques de l'information dans une interface de formation, influe sur la performance psychophysiologique d'un utilisateur. Ce paradigme propose une méthode innovante qui n'avait pas encore été utilisée pour tester la théorie de l'adéquation cognitive (Vessey, 2006; Bacic, 2014).

Implications managériales

D'un point de vue pratique, les résultats de ce mémoire engendrent des implications concrètes pour le secteur de l'aviation. Premièrement, ce mémoire fournit une compréhension claire des facteurs comportementaux qui doivent être surveillés afin d'améliorer les performances des instructeurs de vol en contexte de formation en simulateur. La communauté aéronautique pourrait se fier à ces facteurs afin de développer de meilleures pratiques en matière de sélection, de formation et de développement de l'expertise des instructeurs de vol. Une meilleure compréhension des facteurs de performance pourrait contribuer à améliorer la qualité de l'enseignement des instructeurs, ce qui pourrait conduire à une amélioration de la qualité du processus d'apprentissage des pilotes de vol.

Deuxièmement, pour les spécialistes des facteurs humains en aviation, les résultats du mémoire engendrent plusieurs implications managériales. Les résultats démontrent l'importante de concevoir une interface offrant une meilleure adéquation cognitive entre les représentations et les tâches. Les résultats démontrent également la pertinence d'utiliser différents modèles de visualisations pour évaluer le design des interfaces. Les méthodes d'analyse oculométriques ont permis de tirer des conclusions exploitables sur l'interface utilisée dans l'expérience. À l'aide de l'oculométrie, ce paradigme pourrait être potentiellement être utilisé pour créer une interface qui s'adapte à interaction qu'un utilisateur réalise avec celle-ci. Il sera possible d'utiliser ce paradigme lors de projet futur pour tester l'expérience utilisateur et la conception des interfaces, ce qui aidera le spécialiste des facteurs humains à créer des interfaces plus adaptées aux tâches des instructeurs et pilotes de vol.

Troisièmement, ce mémoire fournit également aux instructeurs des informations sur les facteurs pouvant impacter leur performance. Nos résultats suggèrent que différents modèles visuels envers les représentations de l'information ont une influence sur leur temps d'évaluation et la charge cognitive. Les instructeurs pourraient utiliser le paradigme expérimental proposé dans ce mémoire lors d'une session de formation pour mieux comprendre leur processus attentionnel, en comprendre l'impact et utiliser cette information pour améliorer la manière dont ils interagissent avec leur interface.

Limites du mémoire et pistes de recherches futures

Cette recherche comporte plusieurs limitations. Premièrement, l'échantillon de notre expérience est relativement faible, car ils appartiennent à une population très spécifique, petite et très expérimentée. Ceci a un impact direct avec l'analyse statistique et la généralisation des résultats. En nous basant sur la littérature portant sur la théorie de l'adéquation cognitive, il est étonnant que la majorité de nos hypothèses ne soit pas supportée (Bacic, 2014 ; Vessey, 2006 ; Teets et al., 2010). Cela pourrait s'expliquer par la petite taille de l'échantillon utilisée dans cette expérience, qui ne pourrait pas être suffisante pour supporter nos hypothèses et supporter les études précédentes étudiant la théorie de l'adéquation cognitive. Des études ultérieures pourraient tout de même utiliser le paradigme expérimental développé avec cette étude en augmentant la taille de l'échantillon pour valider les résultats de celle-ci.

Deuxièmement, lors de l'expérience, le stimulus utilisé dans la séquence d'entrainement a été réutilisé plus tard dans le bloc d'atterrissage. Au cours de la séquence d'entrainement, deux participants sur quatre ont estimé que le même stimulus était plus risqué lors de la deuxième séquence de visionnement. En raison du nombre limité d'essais, il est difficile d'établir une relation directe statistiquement significative entre la visualisation du stimulus qui a été vue pour la deuxième fois et une évaluation du risque plus faible. Troisièmement, la mesure du temps utilisé lors de l'expérience de mémoire représente le temps nécessaire que chaque participant a eu recours pour répondre aux trois questions du questionnaire après chaque essai. Cela représente un facteur confondant qui a limité la possibilité de savoir à quelle partie du questionnaire, la différence du temps d'évaluation est imputable à nos résultats. La validité de notre méthodologie pourrait avoir eu une incidence sur le manque de support apporter à nos hypothèses. Des études ultérieures doivent mesurer le temps consacré à la tâche pour juger la performance de l'élève pilote de manière plus précise.

Le paradigme expérimental, dans lequel plusieurs manœuvres d'atterrissage et de décollage ont été montrées aux participants, a induit une grande variabilité au niveau de l'état psychophysiologique des participants. Dans des expériences avec une taille d'échantillon plus grande, cela devrait permettre d'identifier des relations plus significatives entre ces états et les mesures de performance.

Pour conclure, un plus grand nombre de recherches sur les instructeurs de vol en contexte de formation en simulateur doit être réalisé pour obtenir une meilleure compréhension des facteurs psychophysiologiques influençant sur leurs performances et de l'impact de l'interaction humain-machine sur leurs performances. Concernant les avenues de recherche, tel que mentionné précédemment, il serait intéressant de tester le paradigme expérimental développé dans ce mémoire avec un échantillon plus large afin d'en valider les contributions théoriques et implications managériales.

Bibliographie

Alexander, A. L., Wickens, C. D., & Merwin, D. H. (2005). Perspective and coplanar cockpit displays of traffic information: Implications for maneuver choice, flight safety, and mental workload. The International Journal of Aviation Psychology, 15(1), 1-21.

Aragon, C. R., & Hearst, M. A. (2005, April). Improving aviation safety with information visualization: a flight simulation study. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 441-450). ACM.

Bacic, D. (2014). The Role of Cognitive Effort in Decision Performance Using Data Representations:; a Cognitive Fit Perspective.

González Vega, N. (2002). Factors affecting simulator-training effectiveness. Jyväskylä studies in education, psychology and social research, (207).

de Guinea, A. O., Titah, R., & Léger, P. M. (2014). Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. Journal of Management Information Systems, 30(4), 179-210.

Hoepf, M., Middendorf, M., Epling, S., & Galster, S. (2015). Physiological indicators of workload in a remotely piloted aircraft simulation. AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH HUMAN PERFORMANCE WING (711TH)-HUMAN EFFECTIVENESS DIRECTORATE/APPLIED NEUROSCIENCE BRANCH.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

John, T., & Kundisch, D. (2015). Creativity Through Cognitive Fit: Theory and Preliminary Evidence in a Business Model Idea Generation Context. In Wirtschaftsinformatik (pp. 1283-1297).

Landry, S. J. (2009). HUMAN–COMPUTER INTERACTION IN AEROSPACE. Human-Computer Interaction: Designing for Diverse Users and Domains, 197.

Lazure, C., Dumont, L., El Mouderrib, S., Delisle, J-F., Senecal, S., Leger, P.M. (2018). Evaluation of Certified flight instructors' performance factor in a flight simulation training context. Manuscript submitted for publication.

Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-Tracking Measures in Aviation: A Selective Literature Review. The International Journal of Aerospace Psychology, 1-15.

Shen, M., Carswell, M., Santhanam, R., & Bailey, K. (2012). Emergency management information systems: Could decision makers be supported in choosing display formats?. Decision Support Systems, 52(2), 318-330.

Smallman, H. S., John, M. S., Oonk, H. M., & Cowen, M. B. (2001). Information availability in 2D and 3D displays. IEEE Computer Graphics and Applications, 21(5), 51-57.

Socha, Vladimír & Socha, Luboš & Szabo, Stanislav & Hána, Karel & Gazda, J & Kimličková, Monika & Vajdová, Iveta & Madoran, A & Hanakova, Lenka & Němec, Vladimír & Puškáš, Tomáš & Schlenker, Jakub & Rozenberg, Róbert. (2016). Training of Pilots Using Flight Simulator and its Impact on Piloting Precision.

Sohn, S. Y. and Jo, Y. K. (2003). A study on the student pilot's mental workload due to personality types of both instructor and student. Ergonomics, 46(15), pp. 1566–1577.

Sweller, J., Ayres, P. and Kalyuga, S. (2011) Cognitive Load Theory. Springer, New York, NY, 2011.

Olmos, O., Wickens, C. D., & Chudy, A. (2000). Tactical displays for combat awareness: An examination of dimensionality and frame of reference concepts and the application of cognitive engineering. The International Journal of Aviation Psychology, 10(3), 247-271.

Teets, J. M., Tegarden, D. P. and Russell, R. S. (2010). Using cognitive fit theory to evaluate the effectiveness of information visualizations: An example using quality assurance data. IEEE Transactions on Visualization and Computer Graphics, 16(5), pp. 841–853.

The Boeing Company (2018). 2018 Boeing Pilot & Technician Outlook Personnel.

Thomas, M. J. W. (2000). Enhancing Instructional Systems : The Development of a Tool for Evaluating Instructor and Student', Transport, pp. 41–51.

Transport Canada (2016). Flight Test Guide.

Vessey, I. (1991). Cogntive Fit: A Theory-based analysis of the Graphs versus Tables Literature. Decision Science, 22(2), 219–240.

Vessey, I. (2006). The theory of cognitive fit. Human-Computer Interaction and Management Information Systems: Foundations, 141–183.

Vessey, I., & Galletta, D. (1991). Cognitive Fit: An Empirical Study of Information Acquisition. Information System Research, 2(1), 63–84.

Vidulich, M. A., Wickens, C. D., Tsang, P. S., & Flach, J. M. (2010). Information Processing in Aviation. Human Factors in Aviation, 175–215.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. Transportation Human Factors, 2(2), 99-126.